# Electronics for Radiation Detection

Edited by

Krzysztof Iniewski


CRC Press
Taylor & Francis Group

# Electronics for Radiation Detection

# Devices, Circuits, and Systems

**Series Editor**
*Krzysztof Iniewski*
CMOS Emerging Technologies Inc., Vancouver, British Columbia, Canada

**Electronics for Radiation Detection**
*Krzysztof Iniewski*

**Internet Networks: Wired, Wireless, and Optical Technologies**
*Krzysztof Iniewski*

**Semiconductor Radiation Detection Systems**
*Krzysztof Iniewski*

*FORTHCOMING*

**Radio Frequency Integrated Circuit Design**
*Sebastian Magierowski*

**Semiconductors: Integrated Circuit Design for Manufacturability**
*Artur Balasinki*

**Radiation Effects in Semiconductors**
*Krzysztof Iniewski*

**Electrical Solitons: Theory, Applications, and Extensions in High Speed Electronics**
*David Ricketts*

**Integrated Microsystems: Materials, MEMs, Photonics, Bio Interfaces**
*Krzysztof Iniewski*

# Electronics for Radiation Detection

Edited by
Krzysztof Iniewski

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742



Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4398-1649-3 (Ebook-PDF)

# Contents

# Preface

Human beings historically have had short life spans due to infectious diseases, wars, and natural disasters. Life spans have lengthened considerably in the last century, thanks to improvements in hygiene, medicine, and nutrition. The longer life span, however, has led to a dramatic increase in health care costs and increased efforts to deal with chronic diseases. Further progress in medicine and confinement of exploding health care costs can be expected only with advances in technology, in particular for radiation detectors and front-end electronics.

Among all imaging modalities, those based on ionizing radiation are of primary importance. X-ray-based examinations that include mammography, lung imaging, and computed tomography (CT) scans are used routinely in medical offices worldwide. Single photon emission (SPECT) and positron emission topography (PET) are becoming increasingly popular in nuclear medicine applications in hospitals and large medical practices.

In addition to medical imaging, this book also addresses the applications of radiation detection in other areas, particularly in the rapidly growing field of security applications. Luggage scanning, dirty bomb detection, space missions, nuclear plants, and high energy physics experiments are just a few examples of system applications that utilize x-ray and gamma-ray detection.

Despite different principles of operation among those applications, there are numerous commonalities in the signal processing of signals received by radiation detectors: signal amplification, filtering, multiplexing, and analog-to-digital conversion (ADC). These hardware commonalities among imaging techniques merit the inclusion of all related knowledge and know-how into one publication. After all, equipment for radiation detection that encompasses x-ray, CT, nuclear medicine (SPECT/PET), and security/military uses is a several billion-dollar market that offers tremendous opportunities to integrated circuit (IC) designers.

The book is written by top-notch international experts in industry and academia. The intended audience is practicing engineers with some electronics background. The book might also be used as supplementary material in a graduate course curriculum. I sincerely hope that this book will help improve the understanding of radiation detection electronics and stimulate further interest in the development and use of this equipment to benefit us all.

**Krzysztof (Kris) Iniewski**
Vancouver, 2009

# About the Editor

**Dr. Krzysztof (Kris) Iniewski** manages R&D at Redlen Technologies, Inc., a start-up company in British Columbia. He is also an executive director of CMOS Emerging Technologies, Inc. (www.cmoset.com). His research interests are in hardware design for biomedical and networking applications. From 2004 to 2006, he was an associate professor at the Electrical Engineering and Computer Engineering Department of the University of Alberta where he conducted research on low power wireless circuits and systems. During his tenure in Edmonton, he put together a book for CRC Press titled *Wireless Technologies: Circuits, Systems and Devices*.

From 1995 to 2003, he held various technical and management positions with PMC-Sierra. During his tenure, he led the development of a number of VLSI chips used in optical networks. Prior to joining PMC-Sierra, from 1990 to 1994, he was an assistant professor at the University of Toronto's Department of Electrical Engineering and Computer Engineering. Dr. Iniewski has published more than 100 research papers in international journals and conferences. He holds 18 international patents granted in the USA, Canada, France, Germany, and Japan. He received his PhD degree in electronics (honors) from the Warsaw University of Technology, Warsaw, Poland in 1988. Together with Carl McCrosky and Dan Minoli, he is an author of *Data Networks–VLSI and Optical Fibre*, Wiley, 2008. He recently edited *Medical Imaging Electronics*, Wiley, 2009, *VLSI Circuits for Bio-medical Applications*, Artech House, 2008, *Circuits at Nanoscale: Communications, Imaging and Sensing*, CRC Press, 2008, and *Next Generation Networks*, Wiley, 2010. Kris can be reached at kris.iniewski@gmail.com.

# Contributors

**W. C. Barber**
DxRay, Inc.
19355 Business Center Drive, Suite 110
Northridge, California 91324
USA

**Christian Bäumer**
Westdeutsches Protonentherapiezentrum Essen gGmbH
Universitätsklinikum Essen
Hufelandstr 55
D-45147 Essen, Germany

**Dr. Sven Peter Bönisch**
Helmholtz-Zentrum Berlin für Materialien und Energie
Department WTE
Glienicker Straße 100, D-14109
Berlin, Germany

**Vincente Herrero Bosch**
ITACA/Universidad Politecnica de Valencia
Edificio 8G, Camino de Vera s/n
46022 Valencia, Spain

**Francesco Corsi**
Dipartimento di Elettrotecnica ed Elettronica
Politecnico di Bari
Via Orabona 4
70125 Bari, Italy

**Rafal Dlugosz**
Institute of Microtechnology, EPFL
A. L. Breguet
2 CH-2000
Neuchatel, Switzerland

**Rejean Fontaine**
Department of Electrical Engineering and Computer Engineering
University of Sherbrooke
Quebec, Canada

**Maurizio Foresta**
Dipartimento di Elettrotecnica ed Elettronica
Politecnico di Bari
Via Orabona 4
70125 Bari, Italy

**Gene Frantz**
Texas Instruments Incorporated
12203 Southwest Freeway, MS 752
Stafford, TX, 77477
USA

**N. E. Hartsough**
DxRay, Inc.
19355 Business Center Drive, Suite 110
Northridge, California 91324
USA

**Jan S. Iwanczyk**
DxRay, Inc.
19355 Business Center Drive, Suite 110
Northridge, California 91324
USA

**Seong Jong Hong**
Department of Radiological Science
Eulji University
Gyeonggi-do, Korea
and
Institute of Radiation Medicine
Medical Research Center
Seoul National University College of Medicine
Seoul, Korea

**Sachin Junnarkar**
Brookhaven National Laboratory
Building 535
20 Technology Street
Upton, New York 11973
USA

**Jae Sung Lee**
Departments of Nuclear Medicine, Biomedical Sciences, Brain and Cognitive Sciences
Seoul National University, Seoul, Korea
and
Institute of Radiation Medicine
Medical Research Center
Seoul National University College of Medicine
Seoul, Korea

**Christoph Werner Lerche**
ITACA/Universidad Politecnica de Valencia
Edificio 8G, Camino de Vera s/n
46022 Valencia, Spain

**Nail Malakhov**
Interon AS
Nye Vakaas Vei, 78
N-1395 Hvalstad, Norway

**Cristoforo Marzocca**
Dipartimento di Elettrotecnica ed Elettronica
Politecnico di Bari
Via Orabona 4
70125 Bari, Italy

**Sorin Martoiu**
European Organization for Nuclear Research (CERN)
PH/DT Department, Geneva
CERN-CH1211-Geneve 23
Geneva, Switzerland

**Mark Nadeski**
Texas Instruments Incorporated
12203 Southwest Freeway, MS 752
Stafford, TX, 77477
USA

**Einar Nygård**
Interon AS
Nye Vakaas Vei, 78
N-1395 Hvalstad, Norway

**Angelo Rivetti**
Instituto Nazionale di Fisica Nucleare (INFN)
Via Pietro Giuria, 1, 10100
Torino, Italy

**Roger Steadman**
Philips Research Europe
Weisshausstrasse 2
52066 Aachen
Germany

**J. C. Wessel**
Interon AS
Nye Vakaas Vei, 78
N-1395 Hvalstad, Norway

**Hiroshi Uchida**
Central Research Laboratory
Hamamatsu Photonics K.K.
5000 Hirakuchi, Hamakita-ku
Hamamatsu City, Shizuoka Pref.
434-8601, Japan

**Ryoko Yamada**
Central Research Laboratory
Hamamatsu Photonics K.K.
5000 Hirakuchi, Hamakita-ku,
Hamamatsu City, Shizuoka Pref.,
434-8601, Japan

# 1 The Future of Medical Imaging *Understanding Our True Limitations*

*Mark Nadeski and Gene Frantz*

## CONTENTS



## 1.1 INTRODUCTION

There are those who fear that technology is nearly at the physical limitations of our understanding of nature, so where can we possibly go from here? However, technology is not where our limits lie. Integrated circuits (ICs) have always exceeded our ability to fully utilize the capacity they make available to us, and the future will be

no exception. Indeed, technology does not drive innovation: Innovation and human imagination are the forces that drive technology.

## 1.2 WHERE ARE WE GOING?

The broad field of medical imaging has seen some truly spectacular advances in the last half-century that most of us take for granted. Once marvels only in the laboratory, advances such as real-time and Doppler ultrasonography, functional nuclear medicine, computed tomography, magnetic resonance imaging, and interventional angiography have all become available in clinical settings.

It's easy to sit back in wonder at how far the field of medical imaging has come. However, in this chapter we will glimpse the future. Some of this future is quickly taking shape today, though some of it will not arrive for years, if not decades.

Specifically, we'll look at how advances in medical imaging are based on existing technology; how these technologies will provide more capacity and capabilities than we can conceivably exploit; and how they finally lead to the conclusion that the future of medicine is not limited by what we know, but rather by what we can imagine.

Let's begin by looking at the edge of what is real—that wonderful place where ideas are transformed into reality.

### 1.2.1 The EyeCam

For centuries, humanity has dreamed of being able to make the blind see. And, for as long, restoring a person's eyesight has been considered a feat commonly categorized as "a miracle."

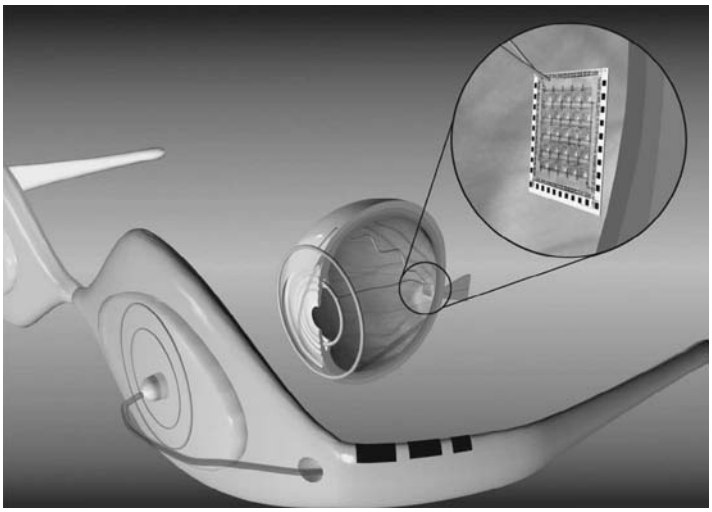


**FIGURE 1.1** Example of the EyeCam, created and tested at the University of Southern California.

About 10 years ago, Texas Instruments (TI) began collaborating with a medical team at Johns Hopkins, well known for its ability to make miracles happen. The team's goal was to develop a way to take the signal from a camera and turn it into electrical impulses that could then be used to excite the retina, as shown in Figure 1.1. If successful, they could return some level of vision back to individuals who had lost their eyesight due to retinitis pigmentosa, a disease that affects more than 100,000 people in the United States alone.

Now at the University of Southern California, this team continues to make significant progress. The project has evolved considerably over the years. Its initial conception consisted of mounting a camera on a pair of glasses that would require patients to rotate their heads in order to look around. Today the team is working to actually implant a camera module *within* the eye, since it is much more natural to let the eye do the moving to point the camera in the right direction. However, it's one thing to say implanting a camera in a person's eye is more practical than mounting it on glasses and quite another to achieve it. A number of challenges come to mind:

*Size*: The complete camera module has to be significantly smaller than an eyeball in order to fit.

*Power*: The camera must have exceptionally low power consumption. At the very least, the energy needs to be scavenged from body heat, the surrounding environment, or a yet-to-be-invented wireless power circuit.

*Heat*: Initial cameras may rely upon a connected power source. Even so, it is critical that the camera not produce much heat. To be practical, the camera must be able to dissipate enough power so as not to heat the eye to the point of discomfort.

*Durability*: The camera must be packaged in such a way as to be protected from the fluids in the eye.

Currently working with Georgia Tech University and experts at TI, the team at USC is busy making all this happen. Is such an ambitious project even possible? Although success has yet to be seen, the team envisions a successful completion of the project. And they have good reason to be confident, for they are only pressing at the edges of possibility.

Much of what lies ahead of us in medicine is the identification of technologies and devices from other parts of our world that we can apply to medical electronics. For example, Prof. Armand R. Tanguay Jr., principal investigator on the EyeCam project, acknowledges that they have many ideas about where else in the body they could implant a camera.*

> Here a camera, there a camera,
> In the eye a little camera.
> Old Doc Donald had a patient.
> E, I, E, I, O.

* Unfortunately, we cannot print any of these exciting ideas without a nondisclosure agreement in place.

Certainly there is more than one verse to this song. The question we might ask ourselves is, "What do we imagine we need next?"

## 1.3 MAKING HEALTH CARE MORE PERSONAL

A device that can help the blind to see is a life-changing application of medical technology. Not all medical devices will have such a dramatic effect on the way we live. Most of the changes in medical care will have a much lower profile, for they will be incorporated into our daily lives. However, while their application may be more subtle, the end result will certainly be profound.

The future of medicine is based upon a firm foundation of existing technologies. What is new, in many cases, is not the technology itself, but rather how the technology is applied in new ways. Consider these key technologies:

- Digital imaging
- Telecommunications
- Automated monitoring

Each of these technologies is already firmly established in a number of disparate industries. Specifically applying them to medicine will still require creativity and hard work, but will enable entirely new applications. Perhaps most importantly, for health care providers and their patients, the resulting advances will help shift health care into becoming a more routine part of daily life, creating a future where medical devices help us to:

1. Manage our chronic conditions
2. Predict our catastrophic diseases
3. Live out our final years in the comfort of our homes

### 1.3.1 Advances in Digital and Medical Imaging

Improving health care is the ultimate goal behind advances in medicine. As medical imaging advances, it will allow patients to have more personalized and targeted health care. Imaging, diagnosis, and treatment plans will continue to become more specialized and customized to a patient's particular needs and anatomy. We may even see therapies that are tailored to genetics. Look at how far we've come already:

*Migration to digital files*: Photographic plates were once used to "catch" X-ray images. These plates gave way to film, which in turn is now giving way to digital radiography. Through the use of advanced digital signal processing, X-ray signals now can be converted to digital images at the point of acquisition while imposing no loss in image clarity. Digital files have a variety of benefits, including eliminating the time and cost of processing film, as well as being a more reliable storage medium that can be transferred near-instantaneously across the world.

*Real-time processing*: The ability to render digital images in real time expands our ability to monitor the body. Using digital X-ray machines during surgical procedures, doctors can view a precise image at the exact time of surgery. Real-time processing also increases what can be done noninvasively. For example, the Israeli company CNOGA* uses video cameras to noninvasively measure vital signs such as blood pressure, pulse rate, blood oxygen level, and carbon dioxide level simply by focusing on the person's skin. Future applications of this technology may lead to identifying biomarkers for diseases such as cancer and chronic obstructive pulmonary disease (COPD).

*Evolution from slow and fuzzy to fast and highly detailed*: Today's magnetic resonance imagers (MRIs) can provide higher quality images in a fraction of the time it took state-of-the-art machines just a few years ago. These digital MRIs are also highly flexible, with the ability to image, for example, the spine while it is in a natural, weight-bearing, standing position. With diffusion MRIs, researchers can use a procedure known as tractography to create brain maps that aid in studying the relationships between disparate brain regions. Functional MRIs, for their part, can rapidly scan the brain to measure signal changes due to changing neural activity. These highly detailed images provide deeper insights into how the brain works—insights that will be used to improve treatment and guide future imaging equipment.

*Moving from diagnostic to therapeutic*: High-intensity focused ultrasound (HIFU) is part of a trend in health care toward reducing the impact of procedures in terms of incision size, recovery time, hospital stays, and infection risk. But unlike many other parts of this trend, such as robot-assisted surgery, HIFU goes a step further to enable procedures currently done invasively to be done noninvasively. Transrectal ultrasound,† for example, destroys prostate cancer cells without damaging healthy, surrounding tissue. HIFU can also be used to cauterize bleeding, making HIFU immensely valuable at disaster sites, accident scenes, and on the battlefield. Focused ultrasound even has a potential role in a wide variety of cosmetic procedures, from melting fat to promoting formation of secondary collagen to eradicate pimples.

*The portability of ultrasound*: Ultrasound equipment continues to become more compact. Cart-based systems increasingly are complemented and/or replaced by portable and even handheld ultrasound machines. Such portability illustrates how, for a wide variety of health care applications, medical technology can bring care to patients instead of forcing them to travel. Portable and handheld ultrasound systems have also been instrumental in bringing health care to rural and remote areas, disaster sites, patient rooms in hospitals, assisted-living facilities, and even ambulances.

---

* www.cnoga.com.

† www.prostate-cancer.org/education/novelthr/Chinn_TransrectalHIFU.html.

*Wireless connectivity*: Portability can be further extended by eliminating cables. Putting a transducer, integrated beam former, and wideband wireless link into an ultrasound probe will not only enable great cost savings by removing expensive cabling from the device, but it will also allow greater flexibility and portability. Further reducing cost and increasing portability enables more widespread use of digital imaging technology, enabling treatment in new areas and applications. A cable-free design also complements three-dimensional (3-D) probes, which have significantly more transducer elements and thus require more cabling, something that may become prohibitively expensive using today's technology.

*The fusion of multiple imaging modalities*: The fusion of multiple imaging modalities—MRI, ultrasound, digital X-ray, PET, and CT—into a single device provides physicians with more real-time information to guide treatment while reducing the time that doctors must spend with patients. For example, positron emission tomography (PET) and computerized tomography (CT) are increasingly being combined into a single device. While the PET scan identifies growing cancer cells, the CT scan provides a picture of the location, size, and shape of cancerous growths.

Many of the real-time imaging modalities have greatly benefitted by advances in digital signal processors (DSP), devices that specialize in efficient, real-time processing. Specifically, the ability to exponentially increase the processing capabilities in imaging machines has enabled these advances to be useful in a hospital setting. However, to drive many of these applications into more widespread usage, another order-of-magnitude increase in processing capability will be necessary. This is a tall challenge.

Fortunately, silicon technology companies are now turning their attention to the world of medical electronics to meet these challenges. For instance, TI formed its Medical Business Unit in 2007 to address the needs of the medical industry. This type of partnership between technology companies and the medical industry will help ensure that the exciting possibilities of the future that we envision will be realized.

### 1.3.2 How Telecommunications Complements Medical Imaging

Advances in medical imaging are frequently complemented by advances in communications networks. Together, they have significantly improved patient care while also reducing costs for health care providers and insurance companies. This development is rapidly moving importance away from where we receive treatment to how we receive treatment.

Telemedicine is the concept where a patient's medical data is transported digitally over the network to a medical professional. For example, 24/7 radiology has begun to emerge as a commonly available service. Instead of maintaining a full radiological staff overnight, a hospital emergency room can now send an X-ray via a broadband Internet link to NightHawk Radiology Services* in Sydney, Australia, or Zurich,

* www.nighthawkrad.net.

Switzerland. NightHawk's staff then reads the X-ray and returns a diagnosis to the ER doctors.

For some patients, the combination of imaging and communications enables diagnosis that they otherwise wouldn't receive for reasons such as finances or distance. A prime example is the work of Dr. Devi Prasad Shetty, a cardiologist who delivers health care via broadband satellite to residents of India's remote, rural villages who otherwise wouldn't receive it simply because of where they live.* Today, one of Dr. Shetty's clinics can handle more than 3,000 X-rays every 24 hours. Shetty's telemedicine program has had a major impact in India, where an average of four people have a heart attack every minute.†

In contrast to telemedicine, telepresence‡ is where a medical professional virtually visits a patient through videoconferencing. Telepresence is increasingly used in both developed and developing countries to widen the distribution of health care. Videoconferencing is often paired with medical imaging systems, such as ultrasound, to enable both telepresence and telemedicine. Such applications frequently enjoy government subsidies because they bring health care to areas where care is expensive, scarce, or both.

One example of telemedicine is the Missouri Telehealth Network,§ whose services include teledermatology. Using this service, a patient at a rural health clinic can put his or her scalp under a video camera for viewing and diagnosis by a dermatologist hundreds of miles away. Videoconferencing equipment simulates a face-to-face meeting, allowing the doctor to discuss any conditions with the patient. In the case of someone with Stage 1 melanoma, early detection via telemedicine may save his or her life.

Whether patients delay a doctor's visit because of distance, cost, available resources, being too busy, or even fear, telemedicine can mean the difference between suffering with a disease or receiving treatment. Virtual house calls, where physicians use videoconferencing and home-based diagnostic equipment to bring health care to a person without necessitating a visit to the doctor, can address most of these concerns.

Virtual house calls may be particularly attractive to patients in rural areas¶ or those in major cities with chronic traffic jams. Virtual house calls are also a way to bring health care to patients who otherwise wouldn't be able to able see a physician, perhaps because they're bedridden, suffering from latrophobia (fear of doctors), or have limited means of transportation. Whatever the hurdle they're helping overcome, virtual house calls are yet another example of how advances in medical technology increasingly are bringing care to the patient instead of the other way around.

For example, let's take a look at how medical technology may have been able to help the 19th-century poet Emily Dickinson, who was reclusive to the point that she would only allow a doctor to examine her from a distance of several feet as she

* www.financialexpress.com/news/Everyone-must-have-access-to-healthcare-facilities-Devi-Prasad-Shetty/42099/.

† www.abc.net.au/foreign/stories/s785987.htm.

‡ www.telepresenceworld.com/ind-medical.php.

§ www.proavmagazine.com/industry-news.asp?sectionID=0&articleID=596571.

¶ www.columbiamissourian.com/stories/2007/05/12/improving-care-rural-diabetics.

walked past an open door. If she were alive today, she would greatly benefit from advances in medical imaging that could accommodate her standoffishness while still diagnosing the Bright's disease that ended her life at age 55.

Future medical technology will reach even further into our lives. Imagine a bathroom mirror equipped with a retinal scanner behind the glass that looks for retinopathy and collects vital signs. In the case of Dickinson, that mirror could have noticed a gradual increase in the puffiness of her face, a symptom of Bright's disease, and alerted her physician through an integrated wireless Internet connection.

One of the underlying technologies behind medical imaging—digital signal processors (DSPs)—has a lot in common with that of telemedicine. DSPs play a key role in telemedicine. For example, DSPs provide the processing power and flexibility necessary to support the variety of codecs used in videoconferencing and telepresence systems. Some of these codecs compress video to the point that a TV-quality image can be transported across low-bandwidth wired or wireless networks, an ability that can extend telemedicine to remote places where the telecom infrastructure has limited bandwidth. In the future, compression will also help extend telemedicine directly to patients' homes over cable and DSL connections.

DSPs also provide the processing power necessary to support the lossless codecs required for medical imaging, since compression could impact image quality and affect a diagnosis. Another advantage DSPs offer is their programmability, which allows them to be upgraded in the field to support new codecs as they become available, thereby providing a degree of future-proofing for hospitals and physicians.

### 1.3.3 Automated Monitoring

Consider this short list of medical devices that can operate noninvasively within our homes:

- Bathroom fixtures with embedded devices could monitor for potential problems, such as a toilet that automatically analyzes urine to identify kidney infections or the progression of chronic conditions such as diabetes and hypertension.
- A bathroom scale could track sudden changes in weight or body fat and then automatically upload this data to a patient's physician. The scale could even trigger scheduling of an appointment based on a physician's predetermined criteria.
- Diagnostic devices such as retinal scanners could be coupled with a patient's existing consumer electronics products, such as a digital camera, to provide additional diagnostic and treatment options. If the device can connect to the network, the medical data collected could automatically be made available to medical personnel.
- Sensors in the home could measure how a person is walking to determine if he or she is at risk for a medical episode such as a seizure.
- Equipment could be connected to a caregiver's network for remote monitoring. One example of such a product under development is a gyroscope-

based device worn by elderly patients to detect whether they have fallen.* Near-falls could trigger alerts to caregivers while being documented and reported to a patient's physician. This device could also track extended sedentary periods, which could be a sign of a developing physical or psychological problem.

- The term *personal area network* (PAN) may come to refer to the variety of devices that work together to regularly and noninvasively monitor and record a person's vital signs. Collected data could be automatically correlated to identify more-complex medical conditions.

All of these examples will change how we approach practical medicine. Passive care of this nature becomes a 'round-the-clock service rather than something that occurs infrequently and disrupts our busy schedules. Constant monitoring also enables earlier identification of health problems before conditions can become irreversible, as well as eliminates the problem of patients not being conscientious about recording information about themselves. As a result, a patient may receive care that's more thorough than if he or she found time for an office visit every week or month.

This technology could be a viable way to improve care for patients who are too busy to schedule routine doctor's visits or, as is the case with latrophobes, whose fear of the doctor has them putting off regular checkups. For the rest of us, who are either too busy, too unconcerned, or too lazy to schedule regular checkups, this technology can help ensure that we don't go too long before any changes get needed care. And, as health care becomes a continuous service through automated processes, the cost of delivering care will be substantially reduced as well.

## 1.4 THE FUTURE OF TECHNOLOGY

This is quite an impressive list of what is just around the corner. And while many of these applications may sound like inventions from a future that we can only hope for and dream about, it is likely that the reality will be even more exciting. To understand why this is the case, let us now shift our attention to the underlying factors that enable and drive innovation.

In the decades since the invention of the transistor, the integrated circuit, the microprocessor, and the DSP, technology has significantly impacted every part of our lives. And during this time, we have seen computers shrink from filling large air-conditioned rooms until they fit into our pockets. Now we're seeing the next stage of this progression as computers move from dedicated devices in our pockets to small subsystems integrated into other systems or even embedded in our clothing or our bodies.

Because it is technology that has brought us where we are, it is tempting to believe that technology is the driving factor behind medical imaging and innovation. However, we don't believe that this is the case. For example, the reason that computers became smaller wasn't because technology allowed them to become smaller. The reason computers became smaller was because people found compelling reasons to

* http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1019448&isnumber=21925.

make them that way. The need or desire for portable computing, not technology, put computers like the Blackberry and iPhone into our pockets. This point is critical: Technology does not drive innovation. Innovation drives technology.

### 1.4.1 Remembering Our Focus

The wants and needs of the marketplace determine the next technology that will go in our pockets as well as the devices we can expect to be embedded in our clothing and bodies. In terms of advances in medical technology, the needs of the patient dictate what comes next.

Technology is exciting, and it's easy to forget what all of these amazing advances are really all about. Whether it's a retinal scanner in a bathroom mirror or a home ultrasound machine, the patient is the greatest beneficiary. These advances enable health care to become more personal by bringing patients to doctors in their offices and doctors to patients in their homes. They also increase the effectiveness of health care by providing ways to identify diseases and other conditions before they become untreatable.

At the same time, these advances also allow health care to fade into the background and become a part of daily life. Imagine being scanned each morning while brushing your teeth instead of only during an annual checkup. That would be particularly valuable for patients with chronic or end-stage diseases, because it may allow them to live their lives without having to move into a hospice.

Health care revolves around the patient, as it should. And technology in turn will help us to address their key needs, as stated earlier in this chapter, to manage their chronic conditions, predict their catastrophic diseases, and allow them to live out their lives in the comfort of their own homes.

### 1.4.2 What We Can Expect from Technology

Knowing that need drives innovation allows us to approach technology from a different perspective, one where it is more relevant to discuss what these advances will be rather than to discuss what will make them possible. For example, we could speculate on the new process technologies that will overcome current integrated circuit (IC) manufacturing limitations. However, this will tell us little about what will be built with these future ICs. By exploring what we can reasonably expect from technology we can gain insight into the possible future.

Let's begin with a well-known tenant from the world of integrated circuits—Moore's law. Moore's law (Moore 1965) forecasts that the number of transistors that can be integrated on one device will double every 2 to 3 years. Made in 1965 by Intel cofounder Gordon E. Moore in 1965, this prediction has not only been adopted by the industry as a "law" but, as shown in Figure 1.2, has accurately described the progress of ICs for the last 40 years.

In practical terms, Moore's law has allowed us to reduce the price of integrated circuits. Advances in performance and power dissipation have also affected cost, but over the last decade advances in IC technology have been primarily responsible for driving cost down.

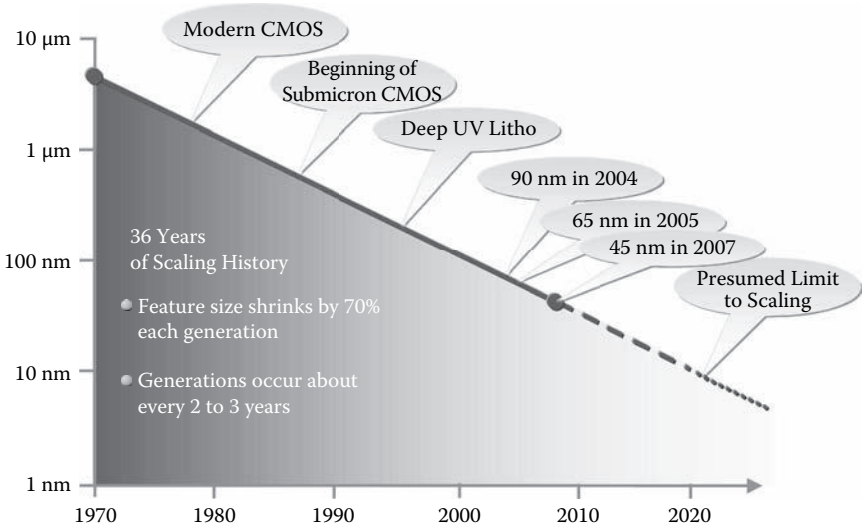



**FIGURE 1.2** The trend for process technology over the last 40 years.

As we look into the future, we can expect Moore's law to continue to hold true; that is, we should be able to integrate about twice the number of transistors on a piece of silicon every 2 to 3 years for about the same cost. What does this mean in terms of medical imaging and innovation? By the year 2020, the price of an IC could be as low as $1 for a billion transistors (a buck a billion). At that time, a high-end processor might cost $50. Just imagine what could be done with 50 billion transistors.

Of course, this many transistors don't come without some issues:

1. The cost of developing a new IC may become prohibitively expensive.
2. Raw performance is no longer driven by Moore's law.
3. Power dissipation must be actively managed.
4. Digital transistors are not particularly friendly to the analog world.

### 1.4.3 Development Cost

Back in the mid-1990s, TI introduced a new technology node for integrated circuits. With this particular new technology node, we were able to integrate 100 million transistors on an IC. To put this in perspective, most personal computers in the 1990s had fewer than 10 million transistors, not counting memory. (As a short aside, advances in new process nodes are generally 0.7 times the size of previous nodes, which yields twice the number of transistors for the same die area.)

Being able to build an IC with this density does not answer the question of what will be done with the transistors. The more difficult question, however, was what it would take to design a product with 100 million transistors. Consider the math. If we estimate that a design effort could be efficient enough such that the average time to design each transistor was about one hour, such a project would take 50,000 staff

years to complete. Clearly, starting a design of 100 million transistors from scratch would be virtually impossible. TI and other IC manufacturers solved and continue to solve this problem through intellectual property (IP) reuse. We'll come back to this concept shortly.

On top of the excessive cost of an IC design, the cost of manufacturing tooling is on the order of $1 million. (We'll overlook the billions of dollars spent on the IC wafer fabrication facility for the moment.) For high-volume applications, this cost spreads out to a manageable number. For example, a design with a total build of 1 million units would reduce the per-unit tooling cost to $1. For low-volume applications, however, the cost can become prohibitive: An IC design for a product anticipated to sell 10,000 units during its lifetime has a per-unit tooling cost of $100.

The obvious conclusion is that only those IC designs with extremely high volumes can be justified. Most applications, however, and not just those in the medical industry, have significantly smaller scope. Thus, these applications will not be based on ICs specifically designed for them but rather on standard programmable processors with application-specific software. And as tooling costs increase, this will be true for virtually all products developed in the future.

Because standard programmable processors allow IP to be implemented in software, the tooling cost of a programmable device can be shared among many different products, even across industries (i.e., medical imaging, high-end consumer cameras, industrial imaging, and so on). Software customizes the processor, so to speak, and the more applications a processor can serve, the lower its cost.

The trade-off of implementing IP in software is, from an engineering perspective, a fairly "sloppy" way of designing a product, meaning that the final design will require far more transistors than if an application-specific IC is used. Given the advanced state of IC process technology, however, this doesn't matter. Back in the 1990s, we had more processors than we knew what to do with. Today we can build processors with more capacity than we can use. And in the year 2020, we will still have more transistors than our imaginations can exploit.

### 1.4.4 Performance

For years, the performance of ICs has seemed to be driven by Moore's law, just as cost was. However, if we measure raw performance—that is, the number of cycles a processor could execute—it actually drifted from Moore's law in the early 1990s. Despite this, processors have still doubled in effective performance in accordance with Moore's law through sophisticated changes to processor architectures, such as deeper pipelines and multiple levels of cache memory. These changes came at their own cost—lots of transistors. Fortunately, as stated before, we have plenty of those.

Improving performance through sophisticated changes to a processor's architecture is, in some ways, just a fancy way of saying that an architecture has been made more efficient. Deeper pipelines, for example, eliminate the inefficiencies of processing a single instruction by simultaneously processing multiple instructions. Eliminating inefficiencies only goes so far, though. Caches improved memory per-

formance significantly, and caches for caches squeezed out a bit more performance, but caches for caches for caches actually slow things down.

There are still many opportunities for increasing processor performance through architectural sophistication, but to achieve a major increase in performance, the industry is moving toward multiprocessing. Also referred to as multicore, the central idea is that, for many applications, two processors can do the job (almost) twice as fast as one. Multiprocessing, while yielding a whole new level of performance, also introduces a whole new level of complexity to processor architectures. And as we continue to add complexity, we then must create more complex development environments to hide the complexity of the architecture from developers.

### 1.4.5 Multiprocessor Complexity

To understand the effect of the complexity that multiprocessing imposes on design, we need to take a look at Amdahl's law (Wikipedia 2010a). Simply stated, Amdahl's law says that sometimes using more processors to solve a problem can actually slow the result. Consider a task such as driving yourself from point A to B. There's really no way to use multiple cars to get yourself there any faster. In fact, using multiple cars along the way will likely slow you down as you stop, switch cars, and then get up to speed again, not to mention the traffic jam you would have created.

The same problem applies to an algorithm that cannot be parallelized—that is, easily distributed across multiple processors. Splitting such an algorithm across equal performance processors will slow overall execution because of the added overhead of breaking the task across the multiple cores. Engineers describe these types of tasks as "Amdahl unfriendly."

Amdahl-friendly tasks are those that can be easily broken into multiple smaller tasks that can be solved in parallel. These are the types of tasks for which DSPs are well suited. Consider how a video signal can be broken into small pieces. Since each piece is relatively independent of the others, the pieces can be processed in parallel simultaneously.

The size of each piece depends upon the processors used. For example, Texas Instruments developed the serial video processor (SVP) in 1994 for the TV market. The SVP contained 1,000 one-bit DSPs, each simultaneously processing one pixel in a horizontal line of video. Because Amdahl-friendly tasks like these can be parallelized, it is easier to determine how to architect the multiprocessing system as well as how to create the development environment with which to design applications that exploit it. See Figure 1.3.

Amdahl-unfriendly tasks are by far much harder to solve, as they prove difficult to divide into parts. Much of the research related to multiprocessing going on now in universities is focused on addressing how to approach such seemingly difficult problems.

#### 1.4.5.1 Multiprocessing Elements

Even given the limitations of Amdahl's law, the best approach for increasing performance appears to be through multiprocessing. Before we can begin to consider how best to take advantage of multiple processors in the same system, however, we

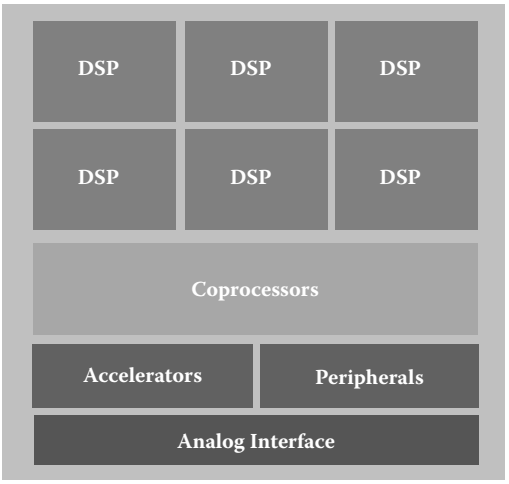



**FIGURE 1.3** Higher performance through parallelism: more DSPs plus flexible coprocessors.

first need to ask, "What is a processor?" There are, after all, many different types of processing elements:

- General-purpose processors: Examples include ARM cores, MIPS cores, and Pentium-class processors.
- Application-specific processors: Examples include DSPs and graphic processing units (GPUs).
- Programmable accelerators: Examples include floating point units (FPUs) and video processors.
- Configurable accelerators: These are similar to programmable accelerators, in that they can perform a range of specific tasks such as filtering or transforms.
- Fixed-function accelerators: These are also similar to programmable accelerators, with the exception that they perform only a single task such as serving as anti-aliasing filters for an audio signal.
- Programmable hardware blocks: Examples include field-programmable gate arrays (FPGAs), programmable logic devices (PLDs), etc.

In general, the term *multiprocessing* refers to heterogeneous (different elements) multiprocessing, while *multicore* refers to homogeneous (same elements) multiprocessing. The importance of this distinction is greater when talking about DSP algorithms than for more general purpose applications. In a DSP application, there is more opportunity to align the various processing elements to the tasks that need to be accomplished. For example, an accelerator designed for audio, video/imaging, or communications can be assigned appropriate tasks to achieve the greatest efficiency. In contrast, very large, generic algorithms may be best implemented using an array of identical processing elements.

Despite all of its challenges, multiprocessing appears to be one of the next major advances that will shape IC, electronics, and medical equipment design. Today, multiprocessing is still not well understood and progress will likely be slow as advances percolate out of university research laboratories into the real world over the next decade or so. In the meantime, we have to get to used to the hit-and-miss nature of multiprocessing architectures and do our best to use them as efficiently as we can.

### 1.4.6 Power Dissipation

In relation to price and performance, power dissipation is the "new kid on the block" and where much research is starting to be focused. TI's first introduction to this important aspect of value goes back to the mid-1950s, when its engineers developed the Regency radio (Wikipedia 2010c) to demonstrate the value of the silicon transistor. This was the first transistor radio, and its obvious need for battery operation made it important to demonstrate low power dissipation.

Power dissipation again became important with the arrival of the calculator in the 1970s. Although most uses for these early calculators allowed for them to be plugged into a wall, sockets were not always nearby or convenient to use. The subsequent movement to LCD calculators with solar cells made low power an even more important requirement.

Lower power dissipation became a primary design constraint in the early 1990s with the arrival of the digital cellular phone. Early customers in this new market made it clear to TI that if power dissipation wasn't taken seriously, they would find another vendor for their components.

With this warning, TI began its now 20-year drive to reduce power dissipation in its processors. One of the results of this reduction in power is the creation of processors that have helped revolutionize the world of ultrasound by turning the once bulky, cart-based ultrasound systems into portable and even handheld systems.

Figure 1.4 shows how power dissipation has improved over the history of DSP development. Measured in units of DSP performance—the MMAC (millions of multiplies and accumulates per second)—it shows how power dissipation has been reduced by half every 18 months. As this chart was created by Gene Frantz, principal fellow at TI, this trend of power efficiency over time has come to be known as Gene's law.

It should be noted that the downward efficiency trend flattened a few years ago. This occurred because of issues with IC technology where leakage power was at parity with active power. As with any problem, once understood it was able to be resolved. Now we are back on the downward trend of power dissipation per unit of performance.

#### 1.4.6.1 Lower Power into the Future

As we look to the future, the question is whether IC technology will continue to follow Gene's law. There are several reasons to believe it will. The two that seem to have the most promise are lower operating voltages and the availability of additional transistors.

Much of the downward trend for power dissipation has been, in fact, due to lowering the operating voltage of ICs. Over the last 20 years, device voltage has gone from

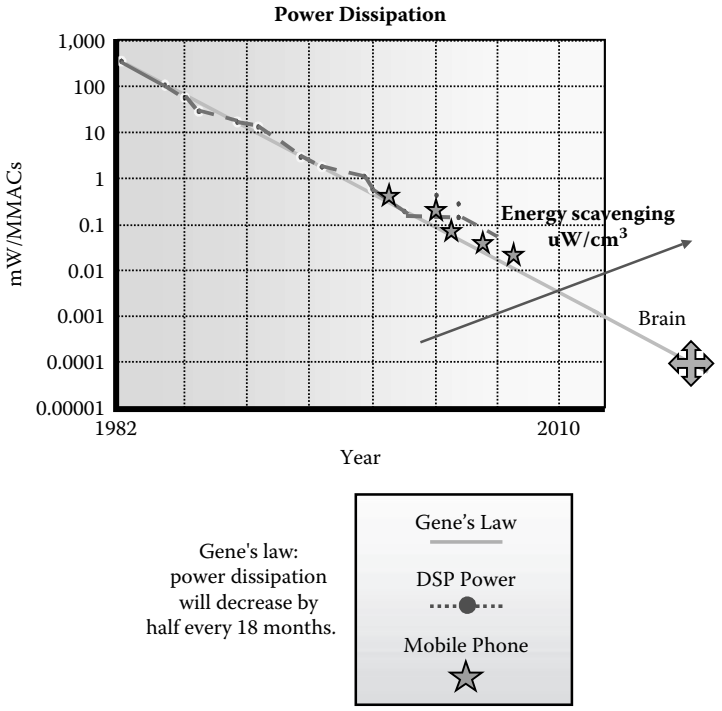



**FIGURE 1.4** Gene's law. This graph shows the trend of power efficiency over time, with efficiency measured as mW/MMAC. The MMAC is the base unit of DSP performance. The upward arrow represents our ability to scavenge energy from the environment.

5 volts to 3 volts to 1 volt and this is not the end of the line. Ongoing research predicts processors operating at 0.5 volts and lower (Chandrakasan et al. 1992).

Just as more transistors can be used to increase the performance of a device, they can also be used to lower the power dissipation for a given function. The simplest method for lowering power consumption is using what is known as *the father's solution*: Who doesn't remember that loud voice resonating through the house, "When you leave a room, turn off the light!"

This wisdom is easily applied to circuit design as well. Simply turn off sections of the device, especially the main processor, when they don't need to be in operation. A good example of this type of management is implemented by the MSP430 family of products (Wikipedia 2010b). And while it does require more transistors to turn only a few sections on and off, this is an extremely efficient approach to power management.

Power can also be managed through multiprocessing, although this approach requires many more transistors. Consider that a task performed on a single processor running at 100 MHz can be performed as quickly on two of the same processors running at 50 MHz. Given the nature of power, one processor running at 100 MHz will consume the same power as two processors running at 50 MHz so long as they are operating at the same voltage.

Due to the characteristics of IC technology, however, the 50-MHz processors can operate at a lower voltage than the 100-MHz processor. Since the power dissipation of a circuit is reduced by the square of the voltage reduction, two 50-MHz processors running at a lower voltage will actually dissipate less power than their 100-MHz equivalent.

The trade-off for managing power through multiprocessing is that two slower processors require twice the number of transistors as a single, faster processor. And again, given that we can rely upon having more transistors as we move into the future, multiprocessing is a feasible method for reducing power dissipation.

#### 1.4.6.2 Perpetual Devices

Confident that IC power dissipation will continue to go down, we can begin to think about how we might take advantage of that. One interesting corresponding area of research that is receiving a lot of attention is energy scavenging. Energy scavenging is based on the concept that there is plenty of environmental energy available to be converted into electrical energy (Shad, Wright, and Rabaey 2003.). Energy can be captured from light, walls vibrating, and variations in temperature, just to name a few examples of sources of small amounts of electrical energy.

Combining energy scavenging with ultralow-powered devices gives us the concept of "perpetual devices." Back in Figure 1.4, there is an upward arrow that represents our ability to scavenge energy from the environment. At the point this line crosses the power-reduction curve, we have the ability to create devices that can scavenge enough energy from the environment to operate without a traditional plug or battery power source.

Imagine the medical applications for perpetual devices. Implants once not feasible because of the need to replace batteries will be possible. Pacemakers will be able to support a wireless link to upload data, eye cams will be permanent once installed, and we may even see roving sensors that travel through our bodies monitoring our heart while cleaning our arteries.

### 1.4.7 Integration through SoC and SiP

So far, we've addressed the three Ps of value: price, performance, and power. The final aspect of IC technology that will serve as a foundation of medical imaging into the future is integration. Integration refers to our ability to implement more functionality onto a single IC as the number of transistors increases. Many in the electronics industry believe that the ultimate result of integration will be what is referred to as an SoC or "system on a chip." There's no nice way to say this: they are wrong.

To understand this position, let's look back at history. When TI began producing calculators, the initial goal was to create a "single-chip calculator." This may be surprising, but such a device has never been produced by TI or anyone else. The reason is that we have never figured out how to integrate a display, keyboard, and batteries onto an IC. What we did develop was a "sub-calculator on a chip." It is important to catch the subtlety here. We didn't develop the whole system, just a part of it. Perhaps it would be more accurate to use the term *subsystems on chip* (SSOC).

The same is true when we look at technology today. No one creates complete systems on chip, and for many good reasons. The best, perhaps, is that in practical

terms, by the time we develop an SoC, we find it has become a subsystem of a larger system. Put another way, once a technology makes sense to implement as an IC (i.e., it has passed the high-volume threshold required to reduce tooling to a reasonable per-unit cost), it has likely been found to be useful in a great variety of applications.

And this leads us to the real focal point of system integration in the future: the system in package (SiP). Figure 1.5 shows that the road map of component integration can be simplified to three nodes. The first node—the design is built on a printed circuit board (PCB)—is well understood by system designers. At the second node, SiP, all of the components are integrated "upward" by stacking multiple ICs into one package. At the third and final node, all of the components are integrated onto a single IC using SoC technology.

Again, once a system warrants its own SoC, invariably it is designed back into larger systems. For this reason, technology isn't stable at the SoC node, and it settles back to placement directly onto a PCB (node 1) or into a package (node 2). As we continue to increase the number of transistors available and lower their cost, these subsystems and the IP they represent will increasingly settle into a package (SiP) and less onto a PCB.

While compelling in itself, this is not the only reason for moving away from the SoC as the ultimate way to design systems. Consider that each advance in digital IC process technology delivers more transistors and perhaps operates at a lower voltage. But the real world is not digital and does not follow this trend, and analog circuits, as well as radio-frequency (RF) circuits, seem to favor higher voltages.

To create a whole IC system on a single piece of silicon requires using a single process technology. However, implementing digital circuits in an analog or RF process significantly increases the relative cost of the digital circuit. Likewise, implementing

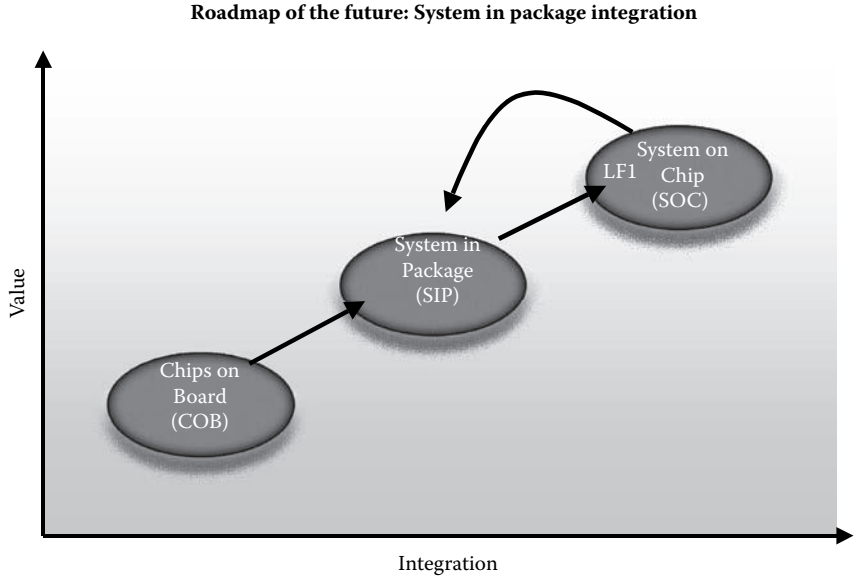



**FIGURE 1.5** A simple road map of component integration.

analog and RF in a digital process substantially reduces signal integrity. The only way we will be able to efficiently integrate all aspects of the system into one "device" is to implement each circuit in its appropriate process technology and combine the various ICs in a single package using SiP technology.

The fact that SiP is more efficient than SoC actually turns out to be really good news, for we will be able to take off-the-shelf devices and stack them in one package that provides virtually all of the advantages an SoC would. The primary difference—and it is significant—is that by using SiP technology we will be able to manufacture devices within months rather than the years required to produce a new SoC design.

And, in much the same way that programmable processors can bring the cost economies of high volume to specialized applications, developers will be able to create highly optimized, application-specific SiPs as if they were standard ICs. This will give rise to a new product concept, the "boutique IC," where system designers can select from a variety of off-the-shelf ICs and "integrate" them into a single package for about the same cost as for the individual ICs. The resulting SiP will have the advantage of faster time-to-market, a smaller footprint, and a "living" specification, where new SoCs can be integrated regularly to continually reduce design costs. This will certainly give new meaning to the concept of "one-stop shopping" for components.

## 1.5 DEFINING THE FUTURE

We've covered a lot of ground in our discussion of the future. When you first read about many of the medical applications we've suggested in this chapter, perhaps you thought they sounded more like science fiction than fact. However, most of these devices build upon existing, proven technologies such as digital imaging, telecommunications, and automated monitoring that will change how we approach medicine.

For the common person, and even latrophobes such as Emily Dickinson, new, noninvasive techniques that are increasingly available in the comfort of their homes will make the difference in diagnosing and treating diseases before they become debilitating or life threatening. For people who live in remote locations, telemedicine will bring doctors and patients together in new and powerful ways.

The underlying IC technology required to bring many of these devices to reality is already, or will soon be, available. Moore's law will continue to give us more transistors than we can conceivably use. More general purpose devices based on software programming models will enable the volumes that result in reasonable cost. Using extra transistors to create multiprocessing circuits will provide higher performance and lower power dissipation. Finally, SiP technology will make possible smaller, more integrated designs as well as potentially enable an entirely new way to design ICs.

As a result, technology is not the limiting factor defining the future of medical imaging. Quite the opposite, technology is more the sandbox in which we can design the creations inspired by our imagination. Innovation is driven not by the fact that we can shrink a computer to fit in a pocket or a body, but rather by the fact that we need or want to shrink that computer.

For the medical industry, the applications that will arise and how fast they will manifest will depend upon what patients need. Technology, for its part, will comply by providing us everything we need to make whatever we envision become real. So what will the future of medical imaging bring? The future will be whatever we want to make it.

The miracles are just beginning.

## REFERENCES

Chandrakasan, A. P., et al. 1992. Low-power CMOS digital design. *IEEE J. Solid-State Circuits* 27:473–484.

Moore, Gordon E. 1965. Cramming more components onto integrated circuits. *Electronics* 38, 8:114–117.

Roundy, Shad, Paul Kenneth Wright, and Jan M. Rabaey. 2003. *Energy scavenging for wireless sensor networks*. Dordrecht: Kluwer Academic Press.

Texas Instruments. 1994. SVP for digital video signal processing. Product overview, SCJ1912.

Wikipedia. 2010a. Amdahl's law. http://en.wikipedia.org/wiki/Amdahl's_law.

Wikipedia. 2010b. TI MSP430. http://en.wikipedia.org/wiki/TI_MSP430.

Wikipedia. 2010c. Transistor radio. http://en.wikipedia.org/wiki/Transistor_radio.

WrongDiagnosis.com. 2010. Statistics by country for retinitis pigmentosa. http://www.wrongdiagnosis.com/r/retinitis_pigmentosa/stats-country.htm.

# 2 Detector Front-End Systems in X-Ray CT
## *From Current-Mode Readout to Photon Counting*

*Roger Steadman and Christian Bäumer*

**CONTENTS**



## 2.1 COMPUTED TOMOGRAPHY

In a computed tomography (CT) scanner, an X-ray source and a detector (normally consisting of many pixel elements) are rotating around a patient. The X-ray beam is absorbed to some extent by bones and soft tissue, forming an intensity profile that is captured by the detector. This intensity profile is normally called a *projection*. As the system rotates, many synchronized projections are taken. Using a back-projection method, the image may be computed for diagnostics [1, 2]. CT detectors differ from other X-ray applications [1, 3] because of their much higher signal resolution and dynamic range (>16 bits). These requirements call for more-sophisticated data-acquisition systems, where noise and dynamic range pose severe difficulties on design issues, as outlined in this chapter.

The technology of medical CT scanners has experienced tremendous progress in the last few years [4]. In particular, there have been significant improvements of the

data acquisition system. For instance, the coverage has been increased through the introduction of 64 slices or more. Further, shorter frame times in data readout make faster scans possible. This, in turn, provides improved diagnostic ability in time-critical applications such as cardiac and other contrast injection studies. Additionally, the increasing power of X-ray tubes necessitated the increase of dynamic signal range. We define the usable dynamic range as the ratio of maximum signal to the double electronic noise floor. Table 2.1 lists the key parameters of state-of-the art CT detectors.

The data acquisition systems described in this chapter belong to the category of solid-state detectors, which can be subdivided into indirect and direct conversion detectors. In indirect conversion detectors, X-rays are stopped in a scintillator, and the converted visible-light photons are transformed into a current in a photodiode array. In CT, $CdWO_4$ crystals and $Gd_2O_2S$:Pr,Ce ceramics (GOS) are commonly used as scintillators. They are structured into individual cells separated by white reflector material such that the cell matrix matches the corresponding photodiode matrix. Pixel width in CT detectors is between 1.0 mm and 1.4 mm. In today's CT detectors, the photodiode array is coupled to an external acquisition system comprising an amplifier and an analog-to-digital converter (ADC). Figure 2.1 sketches a cross section of a detector module [5]. Note that, in a CT scanner, an antiscatter grid, which collimates the X-ray beam to suppress scattered photons, is mounted on top of the detector module. Photodiodes are built in an optically dedicated process optimized for high photoresponsiveness. External electronics normally consist either of discrete components or of custom-made integrated circuits (ICs). Section 2.2 dwells further on the status of data-acquisition systems in current medical CT scanners. Section 2.3 presents alternative electronic readout schemes focusing on

**TABLE 2.1**
**Summary of CT Detector Specifications**

| | |
|---|---|
| Pixel size | $\leq 1.1 \times 1.4$ mm |
| Rise time/fall time | <10 µs |
| Dynamic range | up to 18 bits |
| Frame rate | 1,000 to 10,000 fps |
| Linearity | <0.05% error at any signal level relative to an ideal linear response |
| Number of pixels | $672 \times 128 = 86{,}016$ |
| Maximum energy | 140 keV |
| Maximum X-ray flux | $1\text{–}10 \times 10^9/\text{mm}^2/\text{s}$ |
| Maximum afterglow | <200 ppm after 5 ms |
| | <20 ppm after 500 ms |
| PD[a] responsiveness at 540 nm | ≥0.36 A/W (BIP-based detector) |
| PD[a] responsiveness at 500 nm | ≥0.31 A/W, i.e., quantum efficiency of 0.77 |
| Shunt resistance | ≥1 GΩ |
| Capacitance | <20 pF |
| Crosstalk | <1% |
| Electronic noise floor | 1–5 $pA_{rms}$, depending on frame rate |

[a] PD refers to pixels of the photodiode matrix.

integration of active electronics at pixel level with CMOS (complementary metal-oxide semiconductor) technology.

Semiconductor detectors operated at reverse-bias voltage are employed for direct conversion of X-rays. The mirror charge of the generated charge carriers is sensed at the respective electrodes by amplifiers. Because X-ray photons generate quite fast signals in direct-conversion sensors (on the order of 100 ns), electronic readout can be designed such that stopped X-ray photons are processed individually and counted in a subsequent digital unit. If this detection system is also equipped with electronics for energy discrimination, spectral X-ray detection becomes possible. Currently, there are many research activities that aim to build a spectral X-ray imaging device. Section 2.4 provides some technical background about these detectors and elucidates some aspects of counting-mode detector calibration.

## 2.2 CT DETECTORS TODAY

As visualized in Figure 2.1, the scintillator crystals and photodiodes are held on a large substrate, typically made of ceramic. The electronic unit for signal processing is either located on the substrate (as shown in Figure 2.1) or on a dedicated electronics board, which sits behind the substrate, with crystal and photodiode relative to the X-ray source. A complete CT detector system consists of many detector modules (40 to 60) arranged in an arc centered at the X-ray focal spot.

Detection of optical photons is accomplished by a back-illuminated photodiode (BIP) [5]. Figure 2.2 is a diagram showing a cross-sectional view of the BIP. The bulk silicon of the common cathode device is type $n^-$, with individual anodes for each pixel formed by a $p^+$ diffusion on the lower side. Each pixel anode has a metal pad, which is in turn directly connected to the electronics substrate below (see also Figure 2.1). Light from the scintillating crystal is absorbed at the top of the photodiode, where charge carriers are created. The electrons, being majority carriers in n-type silicon, conduct quite freely to the cathode connection. The holes, being minority carriers in the bulk, must move toward and be absorbed by the anode on the other side before a photocurrent is created. The BIP is operated at zero or low bias. Thus, the bulk of the silicon is field free. This means that holes created by illumination on the backside

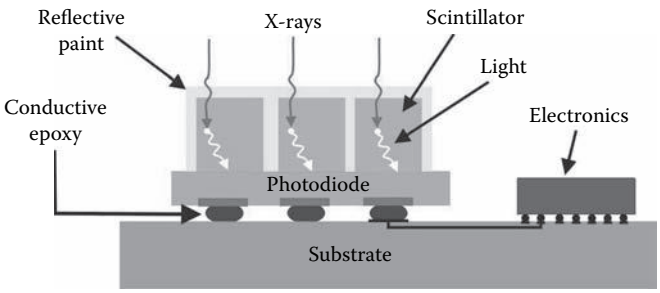



**FIGURE 2.1** Cross section of a CT detector module based on back-illuminated photodiodes. (From *Medical Imaging 2004: Physics of Medical Imaging*, ed. Martin J. Yaffe and Michael J. Flynn, Proceedings of SPIE, vol. 5368 [Bellingham, WA: SPIE, 2004]. With permission.)

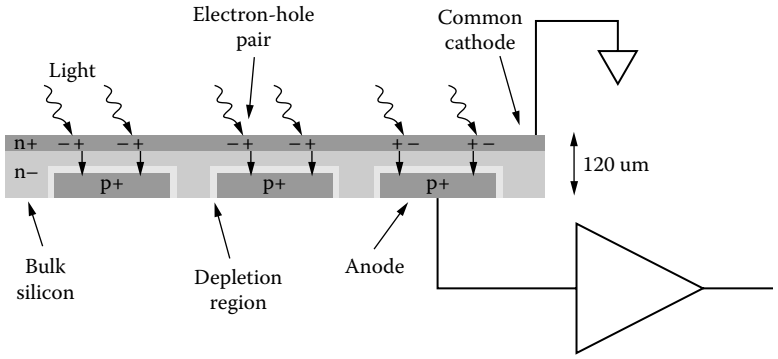



**FIGURE 2.2** Cross section of the back-illuminated photodiode (layer thicknesses not to scale). (From *Medical Imaging 2004: Physics of Medical Imaging*, ed. Martin J. Yaffe and Michael J. Flynn, Proceedings of SPIE, vol. 5368 [Bellingham, WA: SPIE, 2004]. With permission.)

move only by diffusion to reach the anodes. Consequently, hole diffusion in the bulk is the dominant process responsible for crosstalk and the temporal properties of the BIP. With this type of photodiode, rise and fall times on the order of 10 μs can be achieved (see Table 2.1).

Signal processing in the CT detector's front end basically comprises (a) accumulation of the photocurrent of each pixel during frame time and (b) conversion of the obtained charge to a digital signal. The frame time, which is also referred to as the "integration period," can be as short as 100 μs. In current Philips scanners, the photocurrents of the detector matrix are processed by an advanced current-to-frequency converter called TACH ASIC (application-specific IC) [4]. The basic circuit is shown in Figure 2.3. The photodiode current is applied to the input of an integrator. In operation, the integrator output is a voltage ramp with slope proportional to input current. When the integrator output reaches a threshold, a comparator is triggered, which in turn causes a preset amount of charge on a reset capacitor to be dumped at

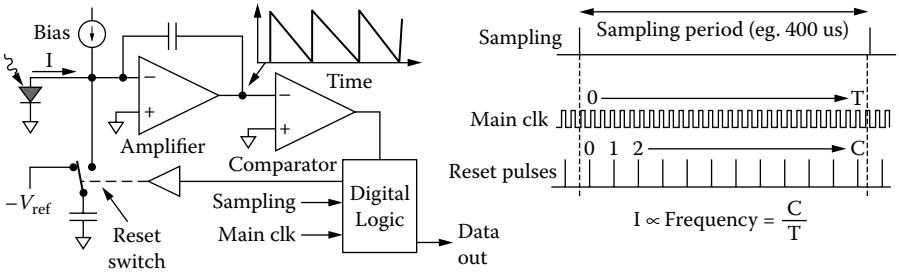



**FIGURE 2.3** Basic circuit and operating principle of the TACH current-to-frequency analog-to-digital converter. A number $C$ of charge packets is counted within the effective frame time $T$ yielding a measure of the photocurrent at the input of the amplifier. (From *Medical Imaging 2006: Physics of Medical Imaging*, ed. Michael J. Flynn and Jiang Hsieh, Proceedings of SPIE, vol. 6142 [Bellingham, WA: SPIE, 2006]. With permission.)

the integrator input. This causes the cycle to start over, and the circuit in effect oscillates at a frequency proportional to the photodiode current at the input. By counting the number of comparator pulses within the frame time along with a measure of the time from the first to last pulse in the frame time, it is possible to get an accurate measurement of the frequency and thus the input current. A small bias current is needed at the input to ensure that at least two pulses occur within the CT sampling period so that a frequency can be determined.

The TACH ASIC can be operated at the required maximum readout rate of 10,000 frames per second (fps). For a frame rate of 2,500 fps, the electronic noise floor is about 2 $pA_{rms}$ [4]. Nonlinearity in the detector is dominated by the electronic readout, since the scintillating crystals and photodiodes are very linear over a wide range. Tests [4] have shown that the nonlinearity of the TACH is well below 0.1%.

## 2.3 CMOS INTEGRATION

Although CT detectors today exhibit excellent performance, bringing the sensor closer to the readout electronics can further improve noise performance and enable quantum-limited operation across the whole dynamic range, i.e., only limited by photon noise. As mentioned previously, today's photodiodes are directly bonded onto the readout electronics through an interposer. Further integration ultimately means bringing the sensor into the same substrate as the electronics. An additional benefit of monolithic integration is the reduction of complexity of the detector and, consequently, a significant cost reduction.

There have been several efforts in the direction of implementing a monolithic CT detector. In a first step, a CMOS photodiode was developed (see Section 2.3.1). Further, an in-pixel current amplifier was developed, which enables the use of CMOS photodiodes in CT detectors today (see Section 2.3.2). Although this is decidedly not a monolithic integration, it already benefits from the introduction of inexpensive CMOS photodiodes in a CT environment. The first reported monolithic CT detector was based on an in-pixel gain-switching amplifier with continuous readout. Details of this implementation and results are presented in Section 2.3.3. Although this ASIC already represents a high degree of integration, the output is still analog, and further analog-to-digital (ADC) conversion is required. The ultimate step is to further integrate the ADC into the same substrate. This is presented in Section 2.3.4, where an in-pixel current-input sigma-delta modulator is used as front-end electronics.

### 2.3.1 CMOS Photodiode

Photodiodes can be simply realized in standard CMOS processes using p–n junctions. Modern downscaled processes, however, typically require substrates exhibiting high doping concentrations and, hence, short diffusion lengths. This fact prevents achieving high responsiveness, and the optimization of it is a rather complicated matter. Low responsiveness is certainly no limitation as long as the signal can be efficiently amplified. Reducing the noise at the input of the amplifier is then of extreme importance for maximum resolution and dynamic range. Other parameters of importance for the application are both dark current and cutoff frequency. Since the amplifier

input capacitance (photodiode + amplifier capacitance) contributes in square-law manner to the noise-power transfer function, the first measure to lower the input equivalent noise is to minimize this capacitance. Additional noise-reduction techniques can be later used in the readout electronics.

A new photodiode structure has been devised. Both low-capacitance and dark-current requirements were addressed by implementing a dot-diode that basically consists of paralleling a number of dot-sized p–n junctions. Figure 2.4 shows a transversal view of the new photodiode. Figure 2.5 visualizes the doping profile as

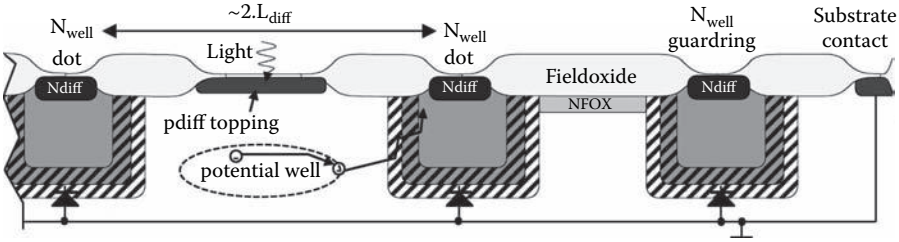



**FIGURE 2.4** Transversal view of the low-capacitance CMOS photodiode with N-dot structuring. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

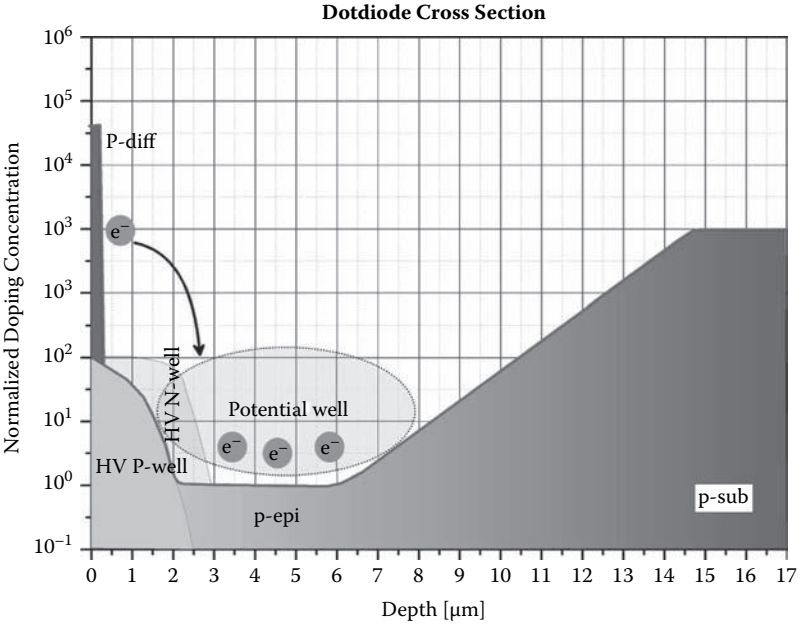



**FIGURE 2.5** Doping profile of the CMOS photodiode across the photon collection area. (From A. Kemna et al., "Low Noise, Large Area CMOS X-ray Image Sensor for CT Application," in *Sensors 3, Proc. IEEE.*, vol. 2, [Piscataway, NJ: IEEE, 2003], 1260–1265. With permission.)

a cross-section along the depth between two dots. Impinging light photons generate electron-hole pairs mainly off the space charge region (on the contrary to area diodes), so that the dominating charge transport is the diffusion of minorities [6].

When using an epitaxial (p-type) substrate, we benefit from having a lower doping concentration near the surface, and thus a potential well is formed. The carrier lifetime and the diffusion length are significantly higher in this region, where minority charges (i.e., electrons) are being collected. During the process, it is important to keep electrons away from the surface to prevent high recombination rates. For this matter, a $p^+$ surface topping ($p_{\text{diff}}$ in Figure 2.4) was added to prevent surface recombination and to push minorities further down into the substrate.

The electrons diffuse then to the $N_{\text{well}}$-dots. The number of dots and distance between them can be optimized for low capacitance at the price of lowering somewhat the responsiveness, and vice versa. A good compromise has been found for a given structure and process. In the present prototype, the distance between dots is 100 μm. This distance is halved at the edge to increase sensitivity and reduce leakage to the guard rings or to neighboring pixels. The way these dots interconnect has been optimized for low capacitance. Redundancy was brought into the design by connecting all dots at both sides and thus lowering the series resistance and improving yield. This structure can be seen in Figure 2.6 (pixel photomicrograph), where the different parts can be identified. Electronics are also depicted in the same figure.

The photodiode size is $1.390 \times 0.830$ mm$^2$. Table 2.2 summarizes the main characteristics of the CMOS photodiode. Although the responsiveness is somewhat lower compared with area diodes and a factor 2.5 to 3 times lower than dedicated photodiodes, designing electronics for low noise benefits enormously from having a capacitance of about 9 times lower (with regard to front-illuminated standard CT

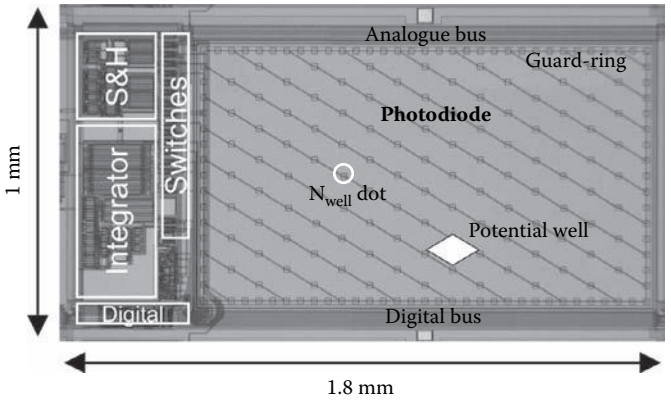



**FIGURE 2.6** Photodiode pixel photomicrograph with corresponding pixel electronics. Electronics correspond to the CMOS monolithic integration, as described in Section 2.3.1. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

**TABLE 2.2**
**Measured Characteristic of CMOS Photodiode**

| | |
|---|---|
| Capacitance | 5.6–6.4 pF [a] |
| Dark current | 150 fA (at 20°C) |
| Responsiveness | 0.13 A/W ($\lambda$ = 480 nm) |
| Cutoff frequency | >15 kHz |

[a] Pixel size is $1.390 \times 0.830$ mm$^2$.

photodiodes). The data of Table 2.2 were obtained at realistic operation conditions, i.e., 1-V reverse bias, and responsiveness was measured at 480 nm, the light output wavelength for the scintillator material used ($CdWO_4$). Trade-offs have to be made on choosing a proper reverse voltage for the photodiode. The higher the reverse voltage, the lower is the capacitance, but dark current increases accordingly. The 1-V reverse bias was considered a good compromise, even though a somewhat higher dark current would not significantly deteriorate overall performance.

### 2.3.2 Current Amplifier

As mentioned previously, the monolithic integration of both photodiode and readout electronics can bring a significant improvement in performance at a lower cost. Computed tomography can, however, already benefit from the introduction of CMOS photodiodes replacing existing optically enhanced arrays. However, this step is not trivial, considering that current CT systems are based on common cathode arrays, whereas CMOS photodiodes are of the common anode type, as outlined in Section 2.3.1. Therefore, the introduction of CMOS photodiodes in CT can only be achieved by either redesigning the front-end electronics or by interfacing both elements in some way. As the CT readout ASICs have been developed over a number of years to optimize noise performance, any topology step requires a long design cycle and multiple verification steps. From a development perspective, designing an interface that allows the use of CMOS photodiodes with existing readout ASIC is preferred. To this end, we propose a current-mode amplifier that not only ensures the right current direction, but also compensates for the lower responsiveness of the CMOS array. Noise performance is critical, as such circuits must not worsen the signal-to-noise ratio (SNR) of the existing readout circuit.

The circuit proposed here is based on a current-mirror approach. Noise is only one aspect that needs to be accounted for. A number of measures also have to be taken into consideration to ensure a good performance in terms of dynamic range, bandwidth, and linearity.

Current-mirror circuits suffer from mismatch in many different forms. It is important to notice, though, that gain errors and current offset do not have a large impact in the application as long as they do not drift. A CT scanner normally goes through thorough calibration procedures to ensure the highest possible image quality. Any contributions that could affect the linearity of the system are indeed

critical. A nonlinearity of more than 0.1% might already cause artifacts in the reconstructed image.

Considering statistics over a number of pixels/chips, two main errors of current mismatch may be identified: the error in the mean μ (or systematic errors) and the error in standard deviation σ (random errors). Systematic errors can be addressed by careful layout of the current-mirror structures, e.g., common-centroid design [7] with large multipliers. The layout has very little influence in the errors in σ, though. These have to be addressed by transistor design and/or circuit measures. Among the most important mismatch mechanisms [7, 8], there are especially two to take care of for the high dynamic range considered here. These are the channel-length modulation and the threshold-voltage mismatch. Considering a simple current mirror with input current $I_1$ and output current $I_2$, the different mismatch phenomena contribute as shown in Equation (2.1), where λ is the channel-length modulation, $V_{GS}$ is the gate-source voltage, and $K$ is $½ \times \mu \times C_{ox}$, with μ being mobility and $C_{ox}$ being gate capacitance. Equation (2.1) assumes the use of large-channel devices. In the present design, the pMOS transistors have a width/length ($W/L$) of 10/15 μm.

$$\frac{I_2}{I_1} = \underbrace{\frac{W_2 \cdot L_1}{W_1 \cdot L_2}}_{\substack{\text{Geometrical} \\ \text{mismatch}}} \cdot \underbrace{\left(\frac{V_{GS} - V_{T_2}}{V_{GS} - V_{T_1}}\right)^2 \cdot \frac{K_2}{K_1}}_{\substack{\text{Voltage threshold} \\ \text{mismatch}}} \cdot \underbrace{\left(\frac{1 + \lambda \cdot V_{DS_2}}{1 + \lambda \cdot V_{DS_1}}\right)}_{\substack{\text{Channel length} \\ \text{modulation mismatch}}} \tag{2.1}$$

Figure 2.7 shows the complete in-pixel electronics, including the primary current mirror consisting of transistors $T_1$ and $T_2$. In our case, in order to minimize the effect of the channel-length modulation, an operational amplifier has been used to clamp the voltage across drain and source ($V_{DS}$) of both pMOS current-mirror transistors. In this way, the circuit is also immune to input offset drift of the readout

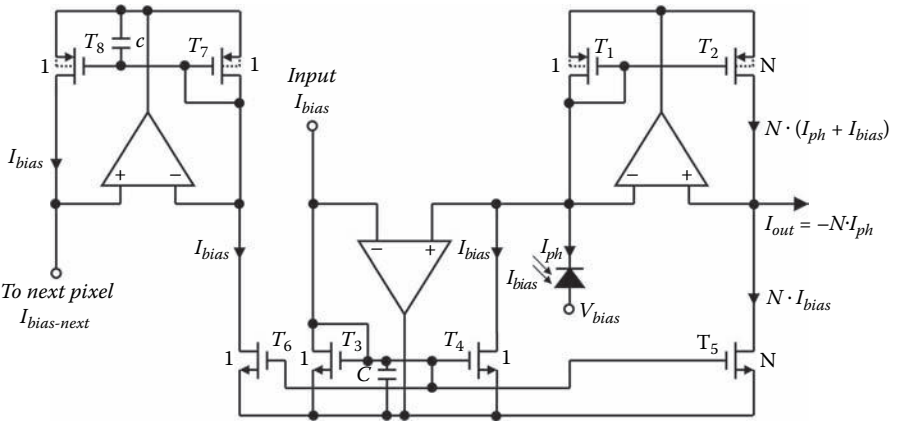



**FIGURE 2.7** Complete in-pixel current-mode amplifier. (From R. Steadman et al., "A High Dynamic Range Current-Mode Amplifier for Computed Tomography," *IEEE Journal of Solid-State Circuits* 41 [2006]: 1615–1619. With permission.)

circuit. Due to linearity requirements and integration issues, a simple single-stage output OTA (operational transconductance amplifier) has been used, the gain of which should be high enough to ensure a low-input offset and a sufficient frequency response. In this design, the open-loop gain is 67 dB, and the gain-bandwidth product, GBW, is 1.5 MHz. The OTA biasing is set to 24 μA. The geometry relation $N$ between both transistors in the current mirror provides the gain of 3 ($N = 3$).

The operation amplifier ensures that $V_{DS1} = V_{DS2}$. However, in nonideal conditions, the OTA has a certain offset that needs to be accounted for to ensure that the nonlinearity contribution by channel modulation can be neglected. In the present design, the maximum OTA offset is in the range of 400 μV. For the considered process and transistor length, λ is approximately 0.02 $V^{-1}$. Taking this into account, the maximum nonlinearity is below 20 ppm. Indirect measurements indicated an OTA offset in the range of 1 mV (corresponding in theory to a maximum nonlinearity of 40 ppm). The readout circuit may also force a certain offset at the output node of the current amplifier and can therefore further contribute to channel modulation. As this offset has been characterized to be below 50 μV, it has been shown that it can be neglected.

The voltage-threshold mismatch is affected by process control, transistor design, and operating-point conditions. It can be demonstrated that, for operation in weak inversion, the voltage-threshold mismatch causes a constant error that does not influence the linearity of the system [9, 10]. As mentioned previously, this would correspond solely to a systematic error and, thus, would not be critical. For the high dynamic range, though, it is not practical to keep the transistor in weak inversion. Since the operating point will certainly move to moderate or even strong inversion, Equation (2.1) needs to be rewritten to consider such conditions, as seen in Equation (2.2).

$$I_2 = \frac{W_2 L_1}{W_1 L_2} \cdot I_1 \cdot \left( 1 - \underbrace{2\sqrt{\frac{\frac{1}{2}\mu C_{Ox} \frac{W_1}{L_1}}{I_1}} \cdot \frac{\overbrace{A_{\Delta V_T}}^{\Delta V_T}}{\sqrt{W_1 L_1}}}_{\text{Distortion}} + \frac{\frac{1}{2}\mu C_{Ox} \frac{W_1}{L_1}}{I_1} \cdot \left( \frac{A_{\Delta V_T}}{\sqrt{W_1 L_1}} \right)^2 \right) \quad (2.2)$$

As can be seen from Equation (2.2), there is a distortion term that depends on the input current $I_1$, and thus it contributes to the nonlinearity. If there were no voltage-threshold mismatch $\Delta V_T$, only a gain error would be present. The actual value of the threshold-voltage mismatch can be determined by the CMOS process characterization. Taking a closer look at Equation (2.2), it can be seen that the best option to minimize the nonlinearity introduced by the threshold-voltage matching is to make the length $L$ as large as possible.

Given the mismatch requirements and noise performance, transistor $T_1$ has to be a large device. For low photocurrents (below 1 nA), the bandwidth requirement can only be fulfilled by biasing the circuit (1 nA). The amplified bias is then subtracted at the output node. The bias is provided by an external source using an nMOS-regulated current mirror of the same kind (transistors $T_3$ and $T_4$). The transistor $T_5$ delivers an amplified $I_{bias}$ (also by a factor $N = 3$). Thus, ideally, no bias will be seen at the output. Due to matching issues, it is very likely that a portion of $I_{bias}$ is still seen at the

output. To a certain extent, as long as it does not compromise the SNR, it is not a problem, since the system undergoes an offset correction before every acquisition. The noise bandwidth is kept low by the capacitors, $C$. Low-frequency noise contributions at the input node will ideally not be seen at the output due to the bias-current subtraction.

Another possible source of nonlinearity is the bulk effect. In the pMOS transistors, the bulks are connected to the source, and thus this effect can be neglected. In the case of the nMOS current-mirror, this connection cannot be done. However, in this case, the current mirror ($T_3$ and $T_4$) sources constant current $I_{\text{bias}}$, and therefore will not be affected by deviations of the bulk-source potential. Only the transistor acting as subtractor ($T_5$) can deviate and cause a nonlinear effect at the output. This is very unlikely, though, since the readout circuit at which the output is connected ensures a virtual ground.

Because CT detectors are typically large-area devices, it is not practical to provide an external bias to every single pixel. A single bias may be provided in parallel to all pixels at the risk of compromising the yield and not having a proper distribution over the chip area. To this end, a so-called bias-regeneration principle has been further integrated at the pixel level. Another pMOS–regulated current mirror ($T_7$ and $T_8$) is used to replicate the input bias. This replicated bias is then fed into the next pixel.

In this way, a number of pixels can be clustered together. Noise bandwidth is also kept low here, thus minimizing the noise propagation through the bias chain. Simulations show that bias deviations of more than ±15% will not significantly disturb the performance in terms of bandwidth. Effectively, more than 16 pixels may be chained together. In reality, though, more pixels may be tied together, since mismatch effects may compensate each other due to statistics. However, bias deviation does have an effect on offset. A changing offset may be seen as a nonlinearity for the low input range. So, accuracy of the bias replica is not important as long as it shows a good stability.

The circuit shown has been realized in a $4 \times 4$ test-structure array fabricated in a 0.8-μm FhG-IMS process. To evaluate whether the pixel clustering has any effect on the performance, all pixels have been chained in different constellations. The supply rails are set at +4 V and –1 V. In this way, the photodiode sees a reverse voltage of –1 V, i.e., a condition that ensures a proper linearity and a good compromise between capacitance and dark current. The power consumption is approximately 350 μW/pixel in dark conditions. Note that the area covered by the active electronics matches roughly the passive area located below the walls of the antiscatter grid and below the septa between the scintillator blocks.

Performance in terms of noise, linearity, and bandwidth was investigated experimentally. Noise has been measured with the proprietary TACH readout circuit intended to be used in combination with the proposed circuit (see Section 2.2). The measured noise amounts to 3 $\text{pA}_{\text{rms}}$ for a time frame of 320 μs (5 $\text{pA}_{\text{rms}}$ at 180 μs). These noise figures correspond to the performance of the readout system, thus indicating that noise is not dominated by the proposed circuit. The same applies for the $1/f$ noise performance.

Linearity was also measured. With the actual test array, a nonlinearity in the range from 0.2% to 2% has been achieved. The voltage-threshold mismatch has been underestimated, yielding significantly worse results than expected. At low current ($I_{bias} >> I_{ph}$), the bias current dominates, i.e., for low current there is an offset mismatch but a fairly linear behavior. The larger the bias current, the better the linearity must be. In this way, it has been possible to confirm that the nonlinearity is caused by an underestimated voltage-threshold mismatch. In order to address this high nonlinearity, trade-offs between a somewhat larger bias, larger transistors, and relaxed noise specification have to be made.

Offset at the output was characterized and found to have a spread of approximately 220 pA over an array. This, however, is not critical for the application. This offset corresponds solely to the mismatch of the bias to the output node. This has been consequently validated by turning off the bias and measuring the offset, which in this case was in the range of 1 pA or less.

Bandwidth has been characterized using a commercial transimpedance amplifier and analyzing the step response. The bandwidth of the current amplifier itself is slightly larger than specified, around 15 kHz, whereas only 10 kHz is needed. A large bandwidth obviously accounts for a higher noise contribution. In principle, this could be tuned by lowering the bias current at the expense of a slightly higher nonlinearity, as mentioned previously.

### 2.3.3 Monolithic Integration of CMOS Photodiode and Readout Electronics

When integrating photodiodes and electronics into the same die, we benefit from a design that is custom made to find the best fit among components. It is often thought that the best single-performance elements yield the best system performance. Nevertheless, this argument is misleading when it comes to in-pixel integration. On the one hand, a standard CMOS process sets certain limitations on the photodiode. On the other hand, electronics will suffer from area constraints, and difficulties arise when considering CT specifications. Thus, we aim at finding the best system performance regardless of the nonideal characteristics of its components [11]. Table 2.1 lists the main specifications for the CT detector. Of main importance are the very high dynamic range and the severe linearity requirements that have to be fulfilled. Such a high dynamic range sets an important challenge in terms of in-pixel integration. In order to build a CT detector with a reasonable geometric DQE (detector quantum efficiency), the area for electronics is constrained to less than 30% of the total pixel size. In this case, a single-stage amplifier is preferable.

A CMOS photodiode was presented in Section 2.3.1, which showed that ultimate performance and cost reduction can be achieved by integrating both sensor and electronics into the same substrate. To this end, we aim at the monolithic integration of both elements into a standard CMOS process.

The CMOS photodiode has been optimized for low capacitance. It can be demonstrated that, for a maximum input equivalent electronic noise current of 0.8 $pA_{rms}$, a maximum photodiode capacitance of 12 pF is allowed (photodiode size

$1.390 \times 0.830$ mm$^2$). The photodiode discussed in Section 2.3.1 fulfills this requirement by showing an equivalent load capacitance of 5.7 pF at −1 V reverse bias.

The electronics consist of an integrator stage with automatic gain switching and a sample-and-hold stage that allows simultaneous integration and readout. This is a mandatory requirement, as in CT it is necessary to maximize X-ray dose usage. The amplifier (integrator stage) has to handle an input current ranging from 6 pA to 173.8 nA. A frame rate higher than 2400 fps and a differential nonlinearity better than 15 bits also constrain the in-pixel electronics design. These limitations have been overcome in the present prototype consisting of a $10 \times 20$ array with a pixel pitch of 1.8 mm $\times$ 1.0 mm$^2$. As described previously, the photodiode-generated current by the scintillator light is integrated over a certain time period. In this section, we will discuss which components and techniques are necessary to obtain a functional channel adhering to specifications emphasizing monolithic integration issues. Figure 2.8 shows the basic topology of the integrator stage.

Due to linearity considerations, an OTA topology is compulsory. High DC gain is necessary to ensure high linearity and low gain error. Mainly due to the area constraints, a folded cascode stage has been implemented. Gain-boosting structures have been necessary to yield an open-loop gain higher than 90 dB. The GBW is higher than 17 MHz. In order to relax specifications in terms of dynamic range, a gain-switching approach has been considered. The idea is to divide the entire dynamic range into two regions so that the OTA performance can be reduced. For the gain ratio of 32 ($2^5$), the OTA dynamic range can be reduced down to 12 bits, while the 5-bit-gain step is indicated by a single-bit digital output used as a gain flag. Gain is set automatically during normal operation, depending on the integrated charge. Figure 2.9 shows the amplifier philosophy where both gain regions can be identified. Such gain-switching is a well-known technique, an example of which may be found in Schanz et al. [12].

The two capacitors in the feedback loop provide the two different gain factors. To implement such a gain-switching approach, an auto-zero comparator is needed to set

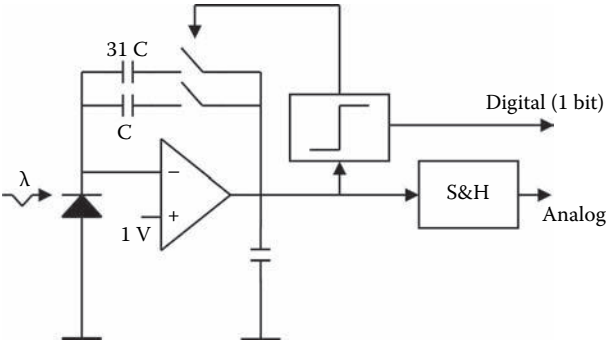



**FIGURE 2.8** Simplified schematic of pixel electronics with automatic gain-switching. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

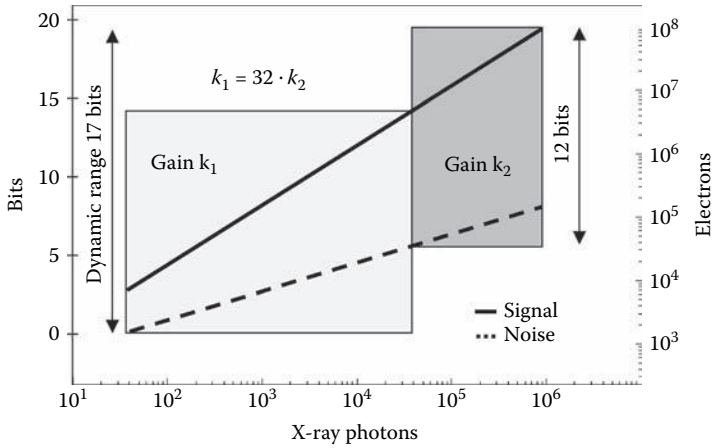



**FIGURE 2.9** Amplifier philosophy, 17-bit dynamic range. Exemplary scaling for an integration period of 400 µs. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

the operating range. During a reset cycle (reset switch not shown in Figure 2.8), both integrator capacitors are discharged to the bias voltage plus the OTA offset,

$$v_C = v_{31C} = 1V + v_{offset}\big|_{I=0} .$$

After that, only the capacitor C is connected, providing feedback to the amplifier, while 31C remains in reset condition. If the integrated charge causes a certain preset threshold voltage to be reached, the comparator trips, which causes the large capacitor 31C to be connected in parallel to C, hence reducing gain by a factor of 32. No charge is lost during the process. For test and characterization purposes the gain switching can be controlled using external signals. The state of the comparator is latched and used as a digital output providing information on the gain status.

Since it is not feasible to read out the 200 pixels during the reset phase, it is necessary to make use of a sample-and-hold circuit (S&H) to enable time-multiplexing of all signals to the output. This S&H block can be seen in Figure 2.10. All pixel outputs will be latched individually (both digital and analog) and read out during the following frame through a single time-multiplexed output. Radiation hardness is not really an issue, since electronics may be placed below the frame for interpixel crosstalk suppression and below the antiscatter grid.

Layout of the integrated circuit is a very important issue when it comes to mixed-signal ICs, and it is particularly critical when considering the very low currents that the input stage has to deal with. Special care has to be taken, and an analysis of capacitive couplings among sensitive nodes (input node mainly) and fast varying signals (digital) has to be made. A parasitic extraction tool was developed to search for any capacitive coupling among nets. It has to be taken into account that a parasitic capacitance of any digital signal to the amplifier input node of just 7.3 aF would cause

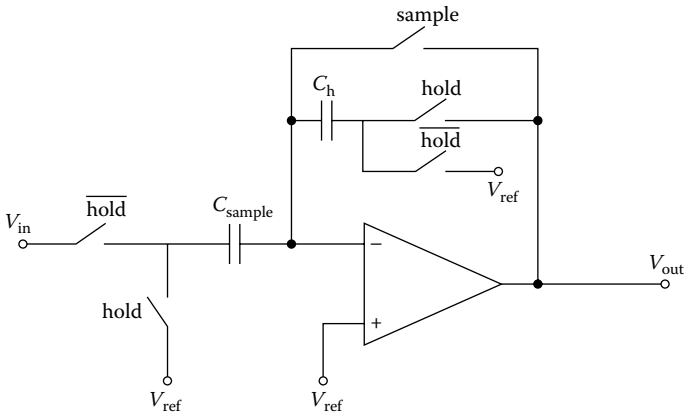



**FIGURE 2.10** Sample and hold (S&H) stage. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

a distortion of about ½ LSB (least significant bit or ADC counts) at 12 bits resolution [6]. Measures were adopted consisting of shielding all sensitive lines, with special attention paid to the integrator input node. Quiet power lines have been used to guard the connecting layer between the photodiode and the integrator input. Total area for electronics is 0.290 mm$^2$. The pixel photomicrograph can be seen in Figure 2.6.

Prior to system implementation, circuit components and design parameters have to be validated by a thorough noise analysis. This study takes into account all noise contributions, including nonideal components like on-resistance of the switches. Noise calculations have to be made for the different topologies that the integrator may present: high-gain, low-gain, and the reset phase. Noise sources are combined, aiming to yield total noise for the two operation modes: high gain and low gain. Throughout this analysis, all noise sources were assumed to be noncorrelated, which is a realistic assumption for the sources themselves and a good approximation for the reset process, provided that it is well characterized.

An indispensable characteristic of a CT detector is that it must be a quantum-limited system, that is to say, all noise sources must combine to a total noise figure that is well below the inherent Poisson noise (often called photon noise) of the incoming signal. This noise analysis aims at finding whether the designed components, combined in a system, do present quantum limited behavior over the CT operating range. Poisson noise can be calculated as the square root of the incoming signal (i.e., shot noise characteristic). We understand the CT operating range as the span of useful X-ray photons that it is aimed at detecting, i.e., from 40 to more than $1 \times 10^9$ X-ray photons/mm$^2$/s. When considering the quantum efficiency of both scintillator and photodiode, a noise specification can be derived. With the selected scintillator and photodiode, the input-referred noise equivalent figure should not exceed the 2000 $e^-$ for quantum-limited operation.

Although the electronic noise is signal independent, the noise at the output of the detector chain is amplified, and thus the high-gain-region performance is of utmost concern (low signal range). The total noise power for every single noise source can be calculated solving:

$$v_n^2 = \int_0^\infty S_{noise} \cdot \left|H_{noise}(f)\right|^2 df, \tag{2.3}$$

where $S_{\text{noise}}$ is the noise-power spectral density, and $H_{\text{noise}}(f)$ is the frequency-dependent noise gain, both depending on the considered noise source [13]. The upper limit of the integral must be chosen according to the detector-chain bandwidth. Even though the S&H stage is considered ideal in terms of noise (it is located after the main amplification stage), the lower cutoff frequency is set by this stage. The noise-equivalent bandwidth is computed then from the 3-dB corner frequency affected by a gain factor, depending on the number of poles (factor 1.22 for a second order low-pass filtering) [13, 14]. For the actual design, the noise-equivalent bandwidth, and, hence, the upper limit for the noise calculation, is 6.6 MHz. Table 2.3 lists the theoretical total figures computed for both high and low gain and the corresponding experimental figures.

The measured electronic noise floor is well below the mentioned specification of 2000 $e^-$. In low gain, noise cannot be evaluated, as the external ADC limits the measurement, as will be described in more detail later in this section. The lower limit was set to 1 Hz to show that the intrinsic 1/*f* noise of the OTA has been suppressed. The reason behind this has to be found on the 1/*f* suppression mechanism [15, 16] taking place during the reset phase, known as correlated double sampling (CDS). During the reset phase, the feedback capacitor is not only discharged to the 1-V bias plus the OTA offset but, in addition, the instantaneous 1/*f* noise is sampled. While integrating, only the difference between the actual values and the sampled one is present at the integrator output. Thus, the shorter the integration time, the lower is the contribution of the 1/*f* noise. In this way, by properly choosing the integration period (generally constrained by system requirements), the 1/*f* contribution may be minimized.

As mentioned previously, all pixels are multiplexed to a single output. A row and a column decoder (four- and five-bit inputs, respectively) have been integrated at chip

**TABLE 2.3**
**Summary of Noise Performance**

| | Theory | | Measured | |
|---|---|---|---|---|
| **Gain Range** | **$e^-$** | **pA [a]** | **$e^-$** | **pA [a]** |
| High gain | 1,085 | 0.43 | 1,000 | 0.4 |
| Low gain | 1,338 | 0.53 | 10,800 [b] | 4.4 [b] |

[a] Integration period = 400 μs.
[b] ADC-limited measurement.

level as well as the column multiplexer. The switches that implement the row multiplexer have been brought into the pixel electronics. A temperature sensor (PTAT, proportional to absolute temperature), with a sensitivity of 1 mV/K attached to a S&H stage with a gain of 3 (output sensitivity 3 mV/K), was also implemented at chip level. The sensor was included to provide information on die temperature and, hence, to enable development of the formal CT calibration and correction algorithms. The sensor is mapped as a pixel and can be read out during scan.

Special care had to be taken when designing and routing the power rails. Since it is indeed a large device and because power consumption is relatively high (700 μA at 5 V, due to single-stage OTA and gain-boosting bias), considerable voltage drop may be present in the supply lines. Since the OTA is sensitive to low supply voltage, a proper layout turned out to be critical.

Both digital and analog outputs are buffered to provide sufficient driving capabilities. The chip output data rate is 576 kHz. The chips have been realized in a FhG-IMS 1.2-μm standard CMOS n-well process, and are three-side buttable, such that two chips can be placed on a single ceramic substrate implementing a 20-slice detector (detector elements in the patient direction). The ASIC size is 2 cm × 2 cm. Figure 2.11 shows an actual CT prototype. One of the chips (bottom one) has a

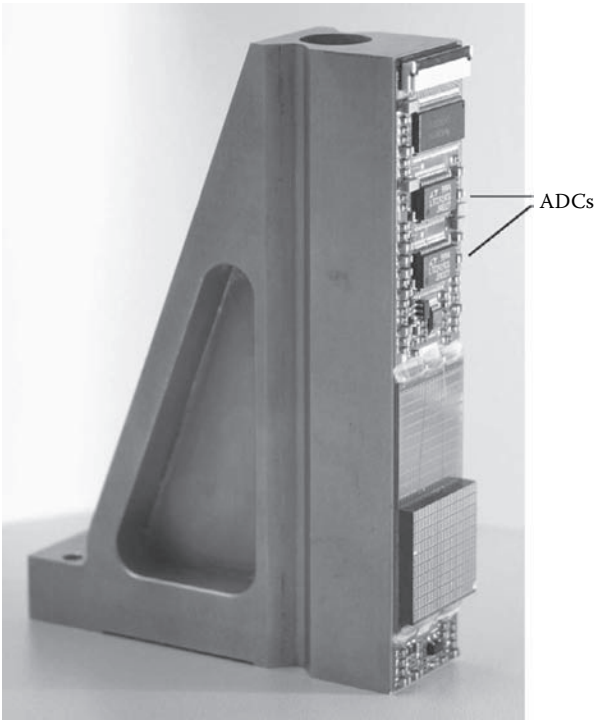



**FIGURE 2.11** Prototype module including two ASICs (three-side buttable). Only one ASIC has a scintillator placed on top. (From R. Steadman et al., "A CMOS Photodiode Array with In-Pixel Data Acquisition System for Computed Tomography," *IEEE Journal of Solid-State Circuits* 39 [2004]: 1034–1043. With permission.)

structured scintillator crystal on top. Electronics on the ceramic substrate comprise two 14-bit ADCs, level shifters, and digital buffers. Reference voltages are derived from the internal ADC reference. Assembly and accuracy issues have been addressed in Spies et al. [17].

A number of characterization procedures have been performed to evaluate the chip performance. Spies et al. [17] present a detailed description on chip characterization. As discussed in Section 2.3.2, important measures of usability for CT detectors are both noise performance and linearity.

Large sets of data measurements have been computed to yield statistics from which noise figures can be derived and compared to the theoretical study. In terms of noise, there are three important figures: temporal noise of single pixels, spatial noise to evaluate for pixel correlations, and 1/*f* corner frequency.

Table 2.3 lists a summary of the noise results for both high and low gain. Results in high gain are in very good agreement with the performance predicted by theoretical calculations. Low gain shows very high noise figures. The reason for this was found not on the chip, but in the external readout electronics. To digitize the analog output, an external 14-bit analog-to-digital converter is used. Generally speaking, the noise floor of such a device is limited to approximately 0.8 LSB. For the considered gain and operating conditions, the chip noise is well below this limit. Consequently, it is not possible to check for noise performance in the low-gain region by simple means. However, this high-noise figure is of no concern, since such level is still well below photon noise for the considered operation range. A more detailed discussion can be found in Steadman et al. [10]. Electronic noise, even when considering the ADC in the low-gain region, is always below the inherent input noise, and thus it can be stated that a quantum-limited system has been achieved. The SNR ranges from 2.6 bits (15.5 dB) at the minimal X-ray dose (or dark conditions) to 9.8 bits (59 dB) at maximum intensity, consistent with quantum-limited operation.

The measurement of 1/*f* consists of evaluating very large nonaveraged data sets, and performing a discrete Fourier transform (DFT) to find pixel-noise spectra. Data sets comprising 200-s acquisitions did not show 1/*f* components. Further measurements were evaluated, and the 1/*f* corner was found to be at about 900 μHz, which is a very interesting result that benefits from the mentioned CDS technique. With a shorter integration period, this corner frequency decreases accordingly. We can conclude that, for the present prototype, 1/*f*-noise components are negligible. When performing a two-dimensional (2-D) spatial DFT, correlations among pixels were not found [17], i.e., crosstalk other than optical or X-rays is negligible.

Linearity is an important issue to prevent artifacts (mainly rings) in the image [18]. It can be demonstrated that image artifacts in CT are proportional to the relative deviation between a linear fit and the sampled data [17]. As mentioned previously, when minimizing the sum of square deviations, the maximum nonlinearity is required to be lower than 0.1%. Integral linearity values have been obtained separately in both high gain and low gain. In the high-gain region, linearity deviations range from 0.4% to 0.9%. In the low-gain region, they were found to vary from 0.6% to 1%. The main causes for this relatively high nonlinearity were found in the output voltage swing of the OTA and its supply rails. The current configuration appears to be particularly sensitive to low supply voltage rails, resulting in a

characteristic bending. Even though this may cause ring artifacts in the projected image, the achieved values are stable enough to be corrected. In CT applications, the whole system undergoes a number of correction and calibration procedures to prevent image artifacts caused by nonuniformities, both gain and offset, and for the nonlinearity that may present. A thorough investigation and characterization on linearity are discussed in Spies et al. [17].

### 2.3.4 Realization of an In-Pixel Sigma-Delta Modulator

Aiming at an even more compact, cost-effective data readout for X-ray CT, a third variant of the CMOS pixel with active electronics is presented. This structure is referred to as *digital pixel*, since each pixel of the matrix encodes its photocurrent into a bit stream that is passed to a common electronics unit. Each sensor element is formed by a photodiode and a third-order sigma-delta (ΣΔ) modulator. The detector has been designed for a bandwidth of 10 kHz, i.e., a sampling frequency $f_s$ of 20 kHz is necessary.

As in any oversampling method, the idea of a ΣΔ converter is to trade resolution in time for resolution in speed. A ΣΔ ADC consists of two parts [19]. In the so-called ΣΔ modulator, the input signal is sampled at high frequency (with oversampling ratio $M$ with respect to $f_s$) and transformed to a pulse density-modulated output signal. The modulator employs the noise-shaping principle, i.e., the quantization noise resulting from the rough quantization through the simple ADC unit, which is often a two-level comparator, is shifted to higher frequencies (modulation noise). This is accomplished by a feedback loop that subtracts a defined signal amplitude from the input node, depending on the result of the last quantization operation. The sequence of outputs from the simple quantizer forms a pulse-density-encoded signal at high speed. It is passed to a filter that down-samples the digital signal to the Nyquist frequency (decimation) and removes the modulation noise with a low-pass filter (LPF) [19].

Similar to the approaches discussed in the previous two sections, a $4 \times 4$ test structure of the detection device has been produced in a FhG-IMS 0.8-µm standard CMOS process [20, 21]. The photodiodes cover 71% of the area. The generic structure of the realized digital pixel is shown in Figure 2.12. Sampling of the photocurrent at elevated rates $M \times f_s = 3$ MHz is performed through the feedback loop of the ΣΔ

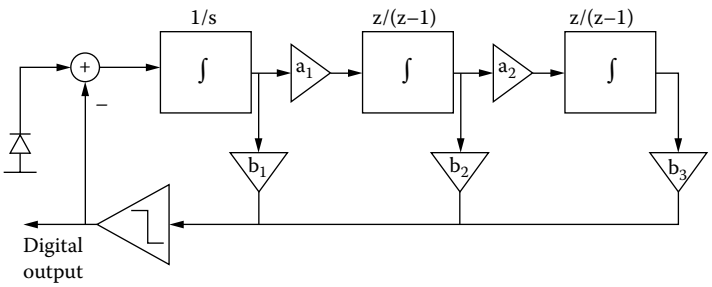



**FIGURE 2.12** Block diagram of the third-order ΣΔ modulator.

modulator. Note that the incoming current, which has negative sign and ranges from 6 pA to 173.8 nA, is directly integrated on the first, continuous-time integrator (1/s). The remaining part of the modulator is in discrete-time operation and driven by two nonoverlapping clocks. A reference voltage $V_m$ acts as mid-voltage for the differential, clocked parts, as reference for the common-mode control of the operational amplifiers and also as reference at the noninverting input of the continuous-time integrator.

Outputs of the continuous-time integrator and discrete-time integrators ($z/[z-1]$) are weighted by the coefficients $a_1$, $a_2$, $b_1$, $b_2$, and $b_3$, and are fed to the comparator. The introduction of weighting coefficients is necessary to ensure stability of the feedback loop. The current feedback (+) is accomplished with a switched-capacitor (SC) source similar to the charge pump of the TACH described in Section 2.2. A 1 at the comparator output initiates the transfer of charge into the first integrator, thereby balancing the incoming flow of electrons on average. In the case of a 0 at the comparator output, the SC current source takes a small amount of charge from the first integrator. Again, this has been implemented to have a stable feedback loop. The SC source is adjusted such that the average bit density is 15% at minimum input signal and 85% at maximum. The overall effect of the feedback loop is a shaping of the quantization noise, i.e., quantization noise is suppressed in the passband at the expense of enhanced quantization noise in the stopband.

The bottom diagram of Figure 2.13 demonstrates the case where the test ASIC has been exposed to light of an LED modulated at 1 kHz. The frequency spectrum exhibits a peak at exactly 1 kHz and some harmonics are attributed to the LED. A linear fit above the signal band from about 10 kHz to about 100 kHz gives a gradient of 40.3 dB/decade. This complies with the expected value of 40.0 dB/decade for

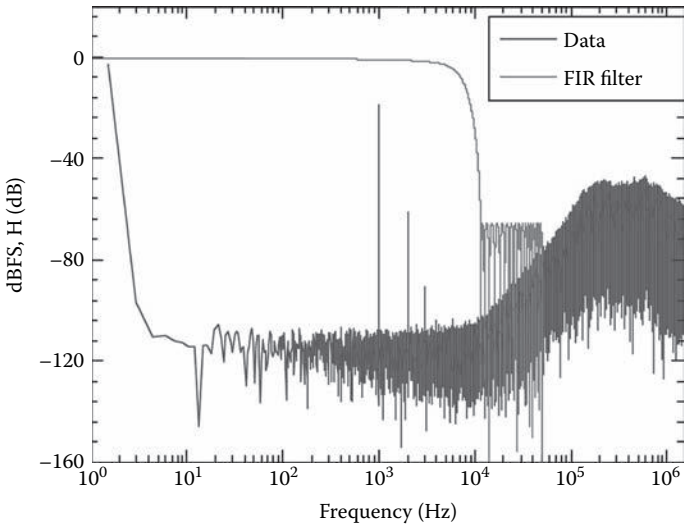



**FIGURE 2.13** Spectrum of a digital pixel where the test structure has been illuminated with a red LED. The spectrum contains $2 \times 10^6$ bins. The transfer function $H$ of the FIR filter, which is the second stage of a decimation filter and operates at 100 kHz, is indicated by the solid line.

a third-order modulator with the used weighting coefficients. The integrated noise power $S_{noise}$ up to 10 kHz yields an equivalent input-referred noise of about 20 $pA_{rms}$. This has to be compared with the specification of a minimum X-ray photon noise of 6.3 $pA_{rms}$. Simulations on the layout level confirm that the noise level is higher than the contributions of modulation noise and thermal noise of the SC source. An additional important noise-coupling path could be identified. It is facilitated by the network of the reference voltage $V_m$. A countermeasure would be to decouple the sensitive input node from the high switching activity of the SC circuitry by establishing a dedicated reference voltage for the first integrator.

According to the principle of ΣΔ conversion, the postprocessing unit has to transform to Nyquist frequency (decimation by a factor $M$) and to apply a LPF that attenuates the modulation noise above the passband (see Figure 2.13). CT imaging requires that the shape of the signal not be distorted in order to preserve the geometrical information in azimuthal direction Φ. This requirement can be met by using finite-impulse response (FIR) filters with linear phase. In particular, sharp transitions in Φ, i.e., in the time domain, have to be reproduced accordingly. It is important to note that this is inherently given for conventional ADCs for X-ray CT (see Sections 2.2, 2.3.1, and 2.3.3), because these integrate the photocurrent over the integration period, which corresponds to a boxcar windowing and, thus, a sync function in the frequency domain. In contrast, many applications of ΣΔ data converters seek to achieve a high resolution over the full bandwidth with sharp cutoff, i.e., a boxcar filter in the frequency domain.

The solution for the described detector is to approximate a sync-filter up to the first zero at 10 kHz. Such a filter can be implemented in a hardware-efficient way employing a two-stage design [19]. First, a sync-filter of third order performs the decimation to an intermediate frequency of 100 kHz. An FIR decimation and LPF with a customized transfer function $H$ as described previously transform from 100 kHz to the Nyquist frequency [21]. The low-pass characteristics of such an FIR filter is indicated in Figure 2.13. Note that the filter reduces the noise current by typically 20%–30%. The filter structure also serves as an anti-aliasing filter on the signal. It could be shown that aliasing streaks emerging at strong transitions from bone to soft tissue are indeed mitigated by this type of digital filter.

## 2.4 Counting-Mode CT Detectors

Photon-counting mode-imaging detectors are regarded as viable options to replace indirect-conversion, integrating-mode devices in many fields of medical X-ray imaging. There are many reasons for a transition to semiconductor radiation detectors, which are read out in single-quanta counting mode. For instance, a higher signal-to-electronic-noise ratio and a better spatial resolution due to a finer pixel segmentation can be achieved. If multiple threshold channels are provided, novel modes of X-ray imaging become possible. Spectral imaging with multithreshold counting detectors appears to be most attractive, as it allows for material separation and quantification [22, 23]. This is especially important in CT, where high-Z elements acting as markers in contrast agents could be identified by the unique imprint of their K-edges on the polychromatic spectrum generated by the X-ray tube [24, 25]. A particular

realization of spectral X-ray imaging is the energy-weighting technique, which yields X-ray images with significantly higher contrast-to-noise ratio compared to acquisitions in integrating mode [26, 27].

A sketch of part of a counting-mode imaging detector is shown in Figure 2.14. The most promising sensor materials for use in X-ray CT are the compound semiconductors CdTe and CdZnTe. For a detective quantum efficiency >98%, which is regarded as necessary in X-ray CT, a sensor thickness of about 3 mm is required. In the simplest configuration, the Cd(Zn)Te crystal is covered by a continuous cathode on its upstream side and by a matrix of rectangular anodes on its downstream side. X-ray photons generate charge carriers within the Cd(Zn)Te, which drift toward their respective electrodes. Readout electronics sense the charge induced on the anodes.

A model for the timing of the induced charged can be derived from the following equation for a single electron [28, 29]:

$$I(t) = e \times \mu_e \times \vec{E}(\vec{x}(t)) \cdot \nabla\varphi(\vec{x}(t)) \tag{2.4}$$

where $e$ denotes the elementary charge; $\mu_e$ the electron mobility; $\vec{x}(t)$ the electron trajectory, with $t$ indicating the time axis; $\vec{E}$ the electrical field; and $\nabla\phi$ the gradient of the anode weighting potential. In more detail, $\varphi(\vec{x}(t))$ describes the ratio of the induced charge on the anode to that of the source charge at position $\vec{x}(t)$. It is a dimensionless value between unity (at the anode) and zero (at all other electrodes) [29].

The trend of the photocurrent is dominated by the component of the weighting field $\partial\phi/\partial z$, which is parallel to the beam direction. Due to a localized weighting

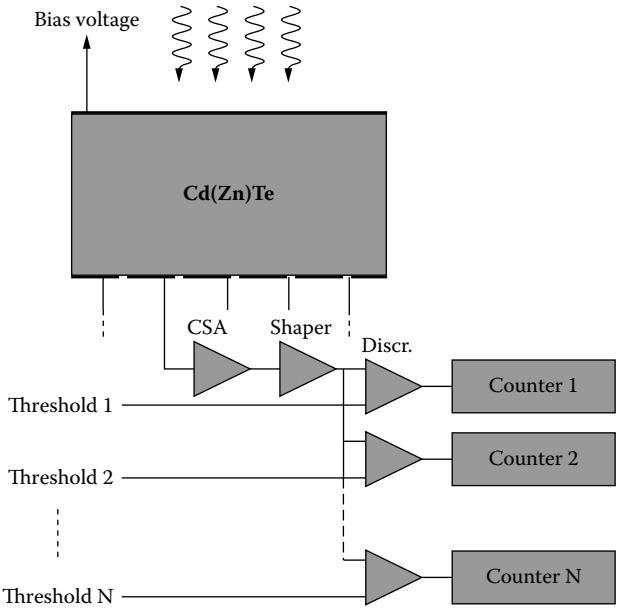



**FIGURE 2.14** Sketch of a direct-conversion, photon-counting detector using Cd(Zn)Te as sensor material.

potential, the electrons produce a signal boost shortly before they arrive at the anode (known as small-pixel effect) [30, 31]. For bias voltages of several hundred volts, the drift time $t_d$ is in the range of 100–200 ns. The boost of the small-pixel effect confines signal induction on the anodes to a time interval of a few tens of nanoseconds. This enables the registration of photon rates of several $10^6$ counts per second per pixel. Fast-counting electronics are necessary to register the current signals induced by single X-ray photon events.

The analog block of each electronic readout channel for counting consists of a preamplifier (or charge-sensitive amplifier, CSA) and circuitry for signal shaping. The shaped signal is passed to $N$ comparators, and the comparator outputs are fed to digital counters (see Figure 2.14). The $N$ energy thresholds of the readout electronics provide a coarse energy quantization. While this data acquisition mode is employed in spectral X-ray imaging, finer energy dispersion is necessary for detector calibration and for detector development. The intrinsic energy resolution of a counting detector can be obtained by scanning at least one discriminator threshold over the range of relevant peak amplitudes. Energy spectra can be derived by computing the derivative.

In contrast to their integrating-mode counterparts, counting-mode detectors have to be characterized and calibrated in the energy domain. Radioactive sources such as $^{241}$Am and $^{57}$Co are readily available to assess the detector response. However, these sources do not cover the energy range of K-edges of possible contrast agents between 25 keV and 55 keV. Further, tests with monochromatic radiation at high flux are not possible. Special beam lines of some synchrotron facilities provide the desired radiation [32, 33].

As an example, energy spectra taken with a detector based on single-line CdTe arrays manufactured by GammaMedica-Ideas (Northridge, California) are presented by Tajima et al. [34]. The detector is equipped with 1024 pixels with 0.4-mm average pixel pitch. The active area of each pixel is $0.38 \times 1.6$ mm with a crystal thickness of 3 mm. The bias voltage was set to 700 V. A lead slit placed in front of the detector crystals reduces the illuminated part of the detector to a height of 1.2 mm.

The experiment was carried out at a beamline of the HASYLAB facility at DESY in Hamburg, Germany. The DORIS-III storage ring contained positrons at 4.44 GeV. Five bunches were filled, yielding typical currents of about 110 mA. The bunches are separated by 192 ns in time. The beamline employs synchrotron radiation that is emitted from a bending magnet of the DORIS-III storage ring. The beam is passed to a double-bounce monochromator equipped with Si(511) crystals. The beam height was about 0.8 mm, and the beam width was about 11 mm. Beam intensity can be changed in two ways: First, by detuning one of the monochromator crystals, the intensity could be reduced. Second, an attenuating metal foil, e.g., made of copper, can be put into the beam path.

Figure 2.15 shows exemplary experimental results for one detector pixel and the scan of one energy threshold measured at a count rate of $10^5$ photons $s^{-1}$. Eight distinct incident energies ranging from 25 to 60 keV were used. The maximum response occurring at the incident energy—the photo peak—can be identified for all incident beams. Electronic noise is the main reason for the smearing in the energy domain. The low-energy tails can be explained by crosstalk effects like, for example,

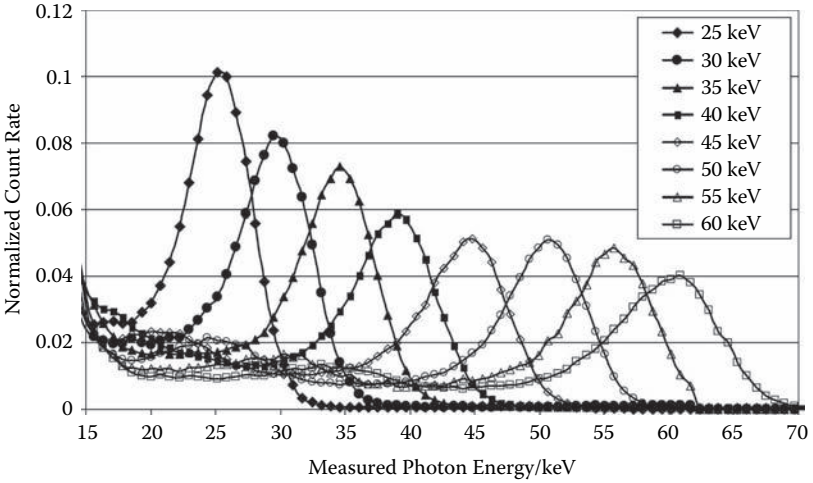



**FIGURE 2.15** Spectral response of one threshold in one detector pixel resulting from monochromatic illumination with photon energies ranging from 25 keV to 60 keV. The count rate in the detector was about $10^5$ photons pixel$^{-1}$ s$^{-1}$. The data are normalized to an integrated count rate of 1 count s$^{-1}$ above 20 keV. (From J. P. Schlomka et al., "Experimental Feasibility of Multi-Energy Photon-Counting K-Edge Imaging in Pre-clinical Computed Tomography," *Phys. Med. Biol.* 53 [2008]: 4031–4047. With permission.)

K-fluorescence and charge diffusion. For instance, a peak is expected at around 25 keV, corresponding to the energy of $K_\alpha$-fluorescence photons from cadmium at 23 keV and tellurium at 27 keV, which originate from X-ray crosstalk from a neighboring pixel. Here, the K-fluorescence photons from cadmium dominate, as they have a larger mean free path ($1/\mu$ = 120 μm for Cd versus 66 μm for Te). Another type of peak is located about 25 keV lower than the photo peak. It is generated by photon absorptions, for which energy is lost due to K-fluorescence (escape peak). Charge diffusion is responsible for the remaining part of the low-energy tails. For a drift time of 100 ns and a diffusion constant of $D$ = 25 cm$^2$/s, one expects a radial spread $\sigma = 22$ μm according to $\sigma = \sqrt{2t_d D}$.

These findings indicate that a compromise in the pixel size must be found; the pixel size should not be too small because of spatial crosstalk. On the other hand, pixel size must be kept small in order to enable a high count rate capability per unit area.

Detector evaluation in the flux domain can also be performed in a synchrotron facility [35]. However, the bunching structure of the synchrotron leads to a time distribution of X-ray photons that deviates from the time distribution of X-ray tubes and radioactive sources. In the experiment at the same facility, a 256-channel line detector, the N-Energy X-ray Imaging Scanning module (NEXIS) [36, 37] of NOVA R&D, Riverside, California, was used. The sensor part is formed by 3-mm-thick CdZnTe arrays with $2 \times 16$ pixels. At the anode side, the height of each row of 16 pixels is 1 mm. The pitch of the anode pixels within each row is 0.964 mm between pixels of the same CdZnTe crystal. A bias voltage of 900 V has been applied. Measurements with a 60-keV monochromatic synchrotron beam have been carried out at different

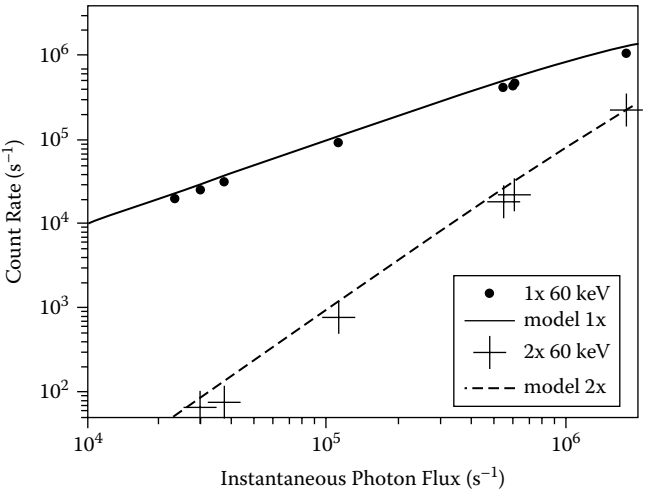



**FIGURE 2.16** Measured event rates for the 60-keV peak and the 120-keV peak as a function of the estimated incident photon rate (monochromatic 60-keV beam). Data points indicate the integral rate of a Gaussian function fitted to the first and second peaks, respectively. The solid and dashed lines represent a model for single- and double-photon events, respectively. (From C. Bäumer et al., "Testing an Energy-Dispersive Counting-Mode Detector with Hard X-Rays from a Synchrotron Source," *IEEE Trans. Nucl. Sci*. 55 [2008]: 1785–1790. With permission.)

settings of the primary intensity. Besides the photo peak at about 60 keV, there is an additional peak at 120 keV. The peak at 120 keV is considered to contain the double-coincidence events. The occurrence of events with photon multiplicity >1 can be used to check the detector behavior with varying X-ray flux and also to perform energy calibration, as will be shown in the following discussion.

In data analysis, the photo peak and the double-coincidence peak have been fitted with Gaussian functions. The integrals of the Gaussian functions are taken to estimate the strength in the respective peaks. The integral count rate of both peaks as a function of the primary X-ray photon flux is shown in Figure 2.16. A direct measurement of the primary X-ray photon flux was not possible. However, the number of counts above the noise threshold can be used to estimate the incident X-ray photon rate, taking into account that a pulsed source was used. Estimation of the photon rate per pixel is outlined as follows.

It is assumed that photons radiated from different bunches can be separated by the detector, and photons generated from the same bunch cannot be resolved [38, 39]. The latter assumption was fulfilled, as the pulse length of a detector pixel of about 350 ns is almost three orders of magnitude larger than the bunch width of about 65 ps. Strictly speaking, the experimental setup does not comply with the former condition of fast detection, because the maximum pulse length exceeds the bunching period $\tau_{\text{bunch}}$ of 192 ns. This deviation from the ideal case will be dealt with later. Assuming that $k$ events above the noise threshold have been registered during a time of $N$ bunch periods, the quantity

$$\hat{p} = \frac{k}{N} \tag{2.5}$$

is an estimate of the probability to register at least one photon per bunch (the present work indicates the estimator of a corresponding parameter by a carat symbol). Since the X-ray photons emitted by a synchrotron bunch follow a Poisson distribution, the mean number of real photons per bunch can be estimated from the count rate measured with an ideal detector by means of the following equation:

$$\hat{r} = -\ln(1-\hat{p}) \tag{2.6}$$

Bateman [40] provides a model for a setup where the dead time of the detector $\tau_{det}$ exceeds the bunch distance $\tau_{bunch}$. Defining $\hat{p}' = k / N$ for the case that a detector with dead time counts $k$ events during $N$ periods, the corresponding parameter $p'$, which denotes the probability of registering at least one event per bunch with such a detector, can be modeled by Equation (2.7) [40]

$$\hat{p}' = \frac{1-e^{-r}}{1+(1-e^{-r})n} \tag{2.7}$$

where $n = \lfloor \tau_{det}/\tau_{bunch} \rfloor$. Thus the real count rate can be estimated from the expression

$$\hat{r} = -\ln\left(1-\frac{\hat{p}'}{1-\hat{p}'n}\right) \tag{2.8}$$

The estimate of the real instantaneous count rate visualized in Figure 2.16 has been calculated by dividing $\hat{r}$ by $\tau_{bunch}$. The number $a$ ($a \in N$) of incident photons per bunch can be measured for any pixel. According to Poisson's law, the mean rate $\lambda_a$ of coincident photons can be described by

$$\lambda_a = \frac{r^a}{a!}e^{-r} \times \tau_{bunch} \tag{2.9}$$

The measured rates of single events and double-coincidence events follow the model of Equation (2.9) in good approximation (see Figure 2.16). When applying Equation (2.9), the estimate $\hat{r}$ from Equation (2.8) has been taken for the rate parameter and $a = 1$ for the 60-keV peak and $a = 2$ for the 120-keV peak, respectively. The model overestimates the singles rate by about 15% over the whole dynamic range. For double coincidences, the model predicts about 30% higher rates than the experimental data points. These deviations are explained by the reduced photo-peak efficiency. For the double-photon events it has also to be considered that the 120-keV peak cannot unambiguously be distinguished from the underlying background, and a conservative estimate of its strength has been applied in the fitting procedure.

To conclude, calibration of photon-counting detectors in the energy domain can be carried out at synchrotron facilities. An important advantage of measurements at

a synchrotron radiation source is that double-photon events (in this case at 120 keV) can be used for energy calibration together with the photo peak of the singles events. This reduces the measurement effort when characterizing X-ray detectors for imaging applications. The respective event rate characteristics are determined by the pulsed synchrotron source, the pulse processing time in each detector pixel, and the photopeak efficiency. Employing an appropriate model, the rate characteristics can be transformed to the Poisson-distributed photon flux encountered with X-ray tubes [38].

Photon counting and especially energy-resolving detectors can be an interesting option for X-ray CT. However, the small pixel sizes necessary for a high-rate capability lead to technical challenges, such as high density of interconnects and readout channels, power consumption, and data rates. The effects of K-fluorescence and charge diffusion have to be observed when going to small pixels. CdZnTe and CdTe are the most promising materials for energy-resolving CT detectors. Good progress has been made to improve the properties of these sensor materials. However, issues with polarization and the behavior under intense X-ray flux remain [32].

However, keep in mind that, for CT applications, the X-ray flux can be as high as ca. $10^9$ quanta/mm$^2$ and that the described type of counting detectors can currently not meet all requirements for CT data acquisition. However, K-edge imaging in preclinical CT with counting-mode detectors has successfully been demonstrated [33]. Moreover, photon-counting-based medical CT for dedicated medical applications has been demonstrated with a research scanner [41].

## 2.5 CONCLUSION

Current CT data-acquisition systems are based on detector modules that comprise a scintillator matrix, a back-illuminated photodiode array, and an advanced current-to-frequency converter for electronic readout. Pursuing an alternative approach for light detection and signal processing, research has proved the feasibility of integrating both photodiodes and electronic readout into a single standard CMOS process. This is a particularly complicated matter when considering the severe requirements that CT detectors have to fulfill. Of main concern was the lower responsiveness of the photodiode, but it has been demonstrated that careful low-noise design and a very low input capacitance can overcome this limitation.

Recent advances in direct-conversion sensors and ASIC design make photon-counting detectors an attractive alternative to integrating-mode detectors. In particular, the energy-discriminating capability of counting detectors allows medical X-ray imaging to enter the spectral domain. This necessitates extensive characterization of the energy-response function of each detector channel. The discussion in this chapter has also demonstrated how a counting-mode detector can be evaluated at a synchrotron facility.

## ACKNOWLEDGMENTS

We thank our colleagues Klaus Jürgen Engel, Michael Overdick, Jens-Peter Schlomka, and Gereon Vogtmeier (Philips Research Europe) and Randy Luhta (Philips Healthcare) for providing some of the material shown in this chapter. The close collaboration with

our colleagues at Philips Healthcare (Israel, the Netherlands, United States) and at Philips Research Europe Aachen/Hamburg is gratefully acknowledged.

## REFERENCES

1. Deych, R., and E. Dolazza. 2007. Advanced X-ray detectors for volumetric CT. In *IEEE Sensors Applications Symposium*, 1–6.
2. Kalender, W. A. 2004. *Computed tomography: Fundamentals, system technology, image quality, applications*. New York: Wiley-VCH.
3. Hsieh, J. 2003. *Computed tomography: Principles, design, artifacts, and recent advances*. Bellingham, WA: SPIE Press.
4. Luhta, R., M. Chappo, B. Harwood, R. Mattson, D. Salk, and C. Vrettos. 2006. A new 2d-tiled detector for multislice CT. *Proc. SPIE* 6142:275–286.
5. Luhta, R., R. Mattson, N. Taneja, P. Bui, and R. Vasbo. 2003. Back-illuminated photodiodes for multislice CT. *Proc. SPIE* 5030:235–245.
6. Kemna, A., W. Brockherde, B. J. Hosticka, E. Oezkan, F. Morales-Serrano, R. Steadman, and G. Vogtmeier. 2003. Low noise, large area CMOS X-ray image sensor for CT application. in *Proc. IEEE Sensors Conf.*, vol. 2, 1260–1265.
7. Pelgrom, M. J. M., H. P.Tuinhout, and M. Vertregt. 1998. Transistor matching in analog CMOS applications. In *Proc. Technical Digest IEDM-98*, 915–918.
8. Pelgrom, M. J. M., A. C. J. Duinmaijer, and A. P. G. Welbers. 1989. Matching properties of MOS transistors. *IEEE Journal of Solid-State Circuits* 24:1433–1440.
9. Steadman, R., G. Vogtmeier, A. Kemna, S. Ibnou Quossai, and B. J. Hosticka. 2005. A high dynamic range, high linearity CMOS current-mode image sensor for computed tomography. In *Proc. European Solid-State Circuit Conference ESSCIRC*, 415–418.
10. Steadman, R., G. Vogtmeier, A. Kemna, S. Ibnou Quossai, and B. J. Hosticka. 2006. A high dynamic range current-mode amplifier for computed tomography. *IEEE Journal of Solid-State Circuits* 41:1615–1619.
11. Steadman, R., F. Morales, G. Vogtmeier, A. Kemna, E. Oezkan, W. Brockherde, and B. J. Hosticka. 2004. A CMOS photodiode array with in-pixel data acquisition system for computed tomography. *IEEE Journal of Solid-State Circuits* 39:1034–1043.
12. Schanz, M., C. Nitta, T. Eckart, B. J. Hosticka, and R. Wertheimer. 1999. A high dynamic range CMOS image sensor for automotive applications. In *Proc. ESSCIRC '99*, 246–249.
13. Gray, P. R., and R. G. Meyer. 1993. *Analysis and design of analog integrated circuits*. 3rd ed. New York: John Wiley & Sons.
14. Ott, H. W. 1988. *Noise reduction techniques in electronic systems*. 2nd ed. New York: John Wiley & Sons.
15. Razivi, B. 2001. *Design of analog CMOS integrated circuits*. New York: McGraw-Hill.
16. Enz, C. C., and G. C. Temes. 1996. Circuit techniques for reducing the effects of Op-Amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization. *Proc. IEEE* 84, no. 11:1585–1614.
17. Spies, L., F. Morales, R. Steadman, K. Fiedler, and N. Conrads. 2003. Performance of prototype modules of a novel multislice CT detector based on CMOS photosensors. *Proc. SPIE* 5030:490–503.
18. Newton, T. H., and D. G. Potts, eds. 1981. *Technical aspects of computed tomography: Radiology of the skull and brain*. Vol. 5. St. Louis: C. V. Mosby Company.
19. Norsworthy, S. R., R. Schreier, and G. C. Temes. 1997. *Delta-sigma data converters: Theory design and simulation*. Hoboken, NJ: John Wiley & Sons.

20. Bäumer, C., M. Gnade, A. Kemna, R. Steadman, G. Vogtmeier, and D. Weiler. 2007. Design and test of an integrated sigma-delta analog-to-digital converter for X-ray computed tomography. *Nucl. Instr. Meth. Phys. Res. A* 576:123–127.
21. Bäumer, C., S. Eick, R. Steadman, G. Vogtmeier, M. Gnade, A. Kemna, and D. Weiler. 2006. Design and evaluation of a CMOS-photosensor with in-pixel sigma-delta modulator for X-ray computed tomography. In *Proc. European Solid-State Circuit Conference ESSCIRC*, 432–435.
22. Alvarez, R., and A. Macovski. 1976. Energy-selective reconstructions in X-ray computerized tomography. *Phys. Med. Biol.* 21:733–744.
23. Giersch, J., M. Firsching, D. Niederlöhner, and G. Anton. 2005. Material reconstruction with spectroscopic pixel X-ray detectors. *Nucl. Instr. Meth. Phys. Res. A* 546:125–130.
24. Riederer, S. J., and C. A. Mistretta. 1977. Selective iodine imaging using K-edge energies in computerized X-ray tomography. *Medical Physics* 4:474–481.
25. Roessl, E., and R. Proksa. 2007. K-edge imaging in X-ray computed tomography using multi-bin photon counting detectors. *Phys. Med. Biol.* 52:4679–4696.
26. Cahn, R., B. Cederström, M. Danielsson, A. Hall, M. Lundqvist, and D. Nygren. 1999. Detective quantum efficiency dependence on X-ray energy weighting in mammography. *Medical Physics* 26:2680–2683.
27. Niederlöhner, D., J. Karg, J. Giersch, and G. Anton. 2005. The energy weighting technique: Measurements and simulations. *Nucl. Instr. Meth. Phys. Res. A* 546 (July):37–41.
28. Shockley, W. 1938. Currents to conductors induced by a moving point charge. *J. Appl. Phys.* 9:635–636.
29. Ramo, S. 1939. Currents induced by electron motion. *Proc. IRE* 27:584–585.
30. Barrett, H. H., J. D. Eskin, and H. B. Barber. 1995. Charge transport in arrays of semiconductor gamma-ray detectors. *Phys. Rev. Lett.* 75:156–159.
31. Eskin, J. D., H. H. Barrett, and H. B. Barber. 1999. Signals induced in semiconductor gamma-ray imaging detectors. *J. Appl. Phys.* 85:647–659.
32. Overdick, M., C. Bäumer, K. Engel, J. Fink, C. Herrmann, H. Krüger, M. Simon, R. Steadman, and G. Zeitler. Status of direct conversion detectors for medical imaging with X-rays. *IEEE Trans. Nucl. Sci*. 56:1800–1809.
33. Schlomka, J. P., E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler, A. Livne, and R. Proksa. 2008. Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography. *Phys. Med. Biol*. 53:4031–4047.
34. Tajima, H., et al. 2004. Performance of a low noise front-end ASIC for Si/CdTe detectors in Compton gamma-ray telescope. *IEEE Trans. Nucl. Sci*. 51:842–847.
35. Bäumer, C., G. Martens, B. Menser, E. Roessl, J. P. Schlomka, R. Steadman, and Günter Zeitler. 2008. Testing an energy-dispersive counting-mode detector with hard X-rays from a synchrotron source. *IEEE Trans. Nucl. Sci.* 55:1785–1790.
36. Cajipe, V. B., R. F. Calderwood, M. Clajus, S. Hayakawa, R. Jayaraman, T. O. Tumer, B. Grattan, and O. Yossifor. 2004. Multi-energy X-ray imaging with linear CZT pixel arrays and integrated electronics. *IEEE Nucl. Sci. Symp. Conf. Rec.* 7:4548–4551.
37. Tumer, T. O., V. B. Cajipe, M. Clajus, S. Hayakawa, and A. Volkovskii. 2006. High-resolution imaging 1D and 2D solid state detector systems. *Proc. SPIE* 6319:63190P.
38. Knoll, G. F. 2000. *Radiation detection and measurement*. New York: Wiley.
39. Bordessoule, M., and M. Lemonnier. 2002. High-rate counters versus integrating devices in synchrotron experiments. *Nucl. Instr. Meth. Phys. Res. A* 481:592–600.
40. Bateman, J. E. 2000. The effect of beam time structure on counting detectors in SRS experiments. *J. Synch. Radiat.* 7:307–312.
41. Iwanczyk, J. S., et al. 2007. Photon-counting energy dispersive detector arrays for X-ray imaging. *IEEE Nuclear Science Symposium Conference Record 2007* 4:2741–2748.

# 3 Photon-Counting Energy-Dispersive Detector Arrays for X-Ray Imaging

*Jan S. Iwanczyk, W. C. Barber, Einar Nygård, Nail Malakhov, N. E. Hartsough, and J.C. Wessel*

## CONTENTS



## 3.1 INTRODUCTION

Medical X-ray imaging is widely used for the diagnosis and treatment of disease. Common diagnostic X-ray imaging applications include mammography, planar radiography, and computed tomography (CT). Mammography is used as a screening

tool for breast cancer detection. Planar radiography is used in a large number of procedures, including the detection of bone fracture in an urgent-care setting and angiography, which is employed before, during, and after coronary catheterization. Computed tomography (CT) is used for the diagnosis of a wide array of conditions, especially in the brain and abdomen, where three-dimensional (3-D) information is needed. Recently, CT has been used for the screening of asymptomatic patients for colon and lung cancer.

In addition to these diagnostic methods, X-ray imaging is becoming increasingly important in the treatment of disease. For example, image-modulated and image-guided radiation therapy (iMRT and iGRT, respectively) use X-ray imaging for treatment planning and during therapy.

Currently, these medical X-ray imaging applications utilize integrating X-ray detectors that sum up the incident X-ray flux over time. Although these integrating systems have been successful, they deliver a large radiation dose to the patient, and they do not have sufficient contrast resolution to discriminate between some tissue types. Recently, we and others have been developing photon-counting energy-dispersive X-ray detector arrays that overcome the limitations of integrating detectors. Significant improvements in patient dose, image quality, and ability to perform tissue discrimination can be achieved with this new technology if requirements for high count rates, good efficiency, and reasonable energy resolution can be met.

We report on the development of an innovative photon-counting detector technology based on cadmium telluride (CdTe) and cadmium zinc telluride (CZT) pixelated arrays electrically connected to application-specific integrated-circuit (ASIC) readout structures for photon counting in X-ray imaging applications. We also present recent results obtained from a photon-counting detector consisting of a silicon photomultiplier (SiPM) coupled to a fast-scintillation crystal. We compare photon-counting semiconductor-based and photon-counting scintillation-based detector performance and utility for X-ray imaging applications.

These new types of X-ray imaging detectors provide fast and highly efficient photon-counting capabilities with sufficiently good energy-dispersive characteristics (energy resolution). Systems based on these new detector technologies may employ optimal energy weighting to reduce X-ray dose to the patient as well as material-decomposition analysis of tissue types to improve contrast. We describe the current state of the art for X-ray imaging detectors, specifically for X-ray CT imaging, and show how photon-counting detectors for X-ray CT can reduce patient dose and provide improved contrast through the use of additional energy information not currently available with integrating detectors.

We review the current development efforts to create photon-counting X-ray imaging detectors and describe the benefits of direct conversion detectors and the trade-offs between candidate detector materials. Our current design of an energy-dispersive photon-counting detector for X-ray CT with CdTe and a readout ASIC is described. Examples of clinical images taken with these CdTe photon-counting detector arrays are presented. We show test results, the ASIC architecture, and the preliminary patient study. We also describe improvements to our detector design that could enhance the count-rate performance. Finally, we describe our work with pho-

ton-counting detectors consisting of SiPMs coupled to fast scintillators, and these results are compared to those from our semiconductor-based detectors.

## 3.2 CONVENTIONAL X-RAY COMPUTED TOMOGRAPHY DETECTORS

The soft-tissue contrast and spatial resolution of images that can be produced by computed tomography (CT) scanners have enabled CT to become one of the most widespread modalities for diagnostic medical imaging. Although used clinically for several decades, CT performance continues to improve as new technologies are incorporated, and so the effort to improve CT detectors is an active area of research. Current CT scanners provide three-dimensional (3-D) images that accurately delineate tissues and help physicians make accurate diagnoses. However, there are two major limitations in conventional integrating CT detector technologies: large patient radiation dose and insufficient contrast resolution to discriminate tissues types.

A CT image is a spatial distribution of the linear attenuation coefficients of an object, μ, where μ at each location is determined by three factors: the chemical composition of the object, the mass density of the object, and the energy of the transmitted X-ray photon.[1] The chemical composition of the object (atomic number $Z$ of each chemical element and its mixture ratio) defines the shape of the curve for attenuation (μ) versus energy, while the mass density (ρ) defines the amplitude (scaling) of the curve. Figure 3.1 shows a curve of attenuation versus energy for three materials. Notice that there is less difference in attenuation coefficients between different materials at higher photon energies in general (circle) and that a specific energy provides the best contrast between materials (arrow).

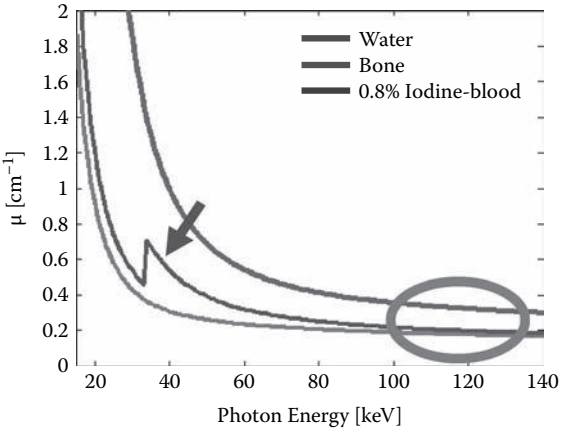



**FIGURE 3.1** Attenuation coefficients for different materials as a function of energy. Notice that the attenuation differences are smaller at higher energy. The top curve represents water, the middle curve bone, and the bottom curve iodine-blood. (Graph courtesy of Katsuyuki Taguchi at Johns Hopkins University.)

The attenuation coefficients of materials are larger in general at lower photon energies. Thus, X-ray photons with lower energy are more heavily attenuated than ones with higher energy. With conventional intensity-integrating CT detectors, the X-ray photon flux—which is transmitted through the patient and is incident on a detector—is converted into a digital signal. The conversion process applies a weight proportional to the photon energy imparted to the detector and adds dark-current noise and Swank noise.[2,3] Photon-counting detectors, however, count each individual photon and therefore do not add extra weight to the higher-energy X-rays. Since these X-rays carry less tissue contrast information as compared to lower-energy X-rays (due to the attenuation difference), this is a benefit of photon counting.

Conventional CT systems use an intensity-integrating detector typically comprising Si photodiodes that are optically coupled to a scintillator. The need to obtain the spatial resolution and soft-tissue contrast that is desired in reconstructed images places high demands on the intrinsic spatial resolution and dynamic range on the detector. To meet these needs, the typical size of pixels in CT detectors is about 1 $mm^2$, and as there can be a tremendous flux of X-rays delivered to the detectors (up to $100 \times 10^6$ photons/$mm^2$/s in air outside the patient region), a large dynamic range with linear response is required.

The detector is typically operated in an integrating mode, where the detector generates a signal proportional to the total energy deposited as a function of time. Each incident X-ray photon undergoes an interaction with a scintillation converter that emits visible light. This light reaches the photodiode component, which provides an electrical signal (current) proportional to the energy flux of X-rays. Since the current is monitored, these detectors are often said to be operating in "current mode." Current-mode readout integrates both the signal and the noise from the detector and electronics over time, and the X-rays in CT have a broad spectrum of energies. When the X-ray flux or the X-ray energies are small, the signal is detected only if it exceeds a noise level produced by the detector and readout electronics. Thus there is a minimum threshold of X-ray flux that can be reliably detected, a threshold that increases as the X-ray energy decreases.

In clinical CT scanners, the incident X-ray beam is highly filtered to remove low-energy photons that cannot be detected, thus reducing patient dose. Therefore, in conventional current-mode CT detectors, there is a distinct nonzero lower limit on the dynamic range. This lower limit of detection defined by noise in the detector causes a pixel with no incident radiation to produce some signal due to electronic noise of the detector and processing electronics. The noise limit in clinical whole-body CT scanners is currently between 0.1% and 0.3%, which places a limit on soft-tissue contrast.[4]

There are additional major deficiencies of integrating (current-mode) systems, such as not taking advantage of statistical information carried by each photon (e.g., three photons of 30 keV carry the same information as one photon of 90 keV) and not using information about the energy of the counted photons. For example, because a polyenergetic X-ray spectrum is used, each detected photon also contributes different information to the resulting image, depending on density and elemental composition of the examined tissue. Utilization of the energy information carried by individual X-ray photons can lead to further improvement of the quality of the image

and/or reduction of the radiation dose. This is accomplished by an optimal energy weighting of low- and high-energy projection images to increase soft-tissue contrast in reconstructed CT images.

Despite these limitations, current-mode X-ray detectors are used in virtually all clinical X-ray systems, including CT, digital radiography, and mammography. Because of these limits, inherent in current-mode detectors, there has been a mounting effort to move away from this technique and explore photon-counting detectors for these X-ray imaging applications.

## 3.3 LOWER DOSE IN CT WITH PHOTON-COUNTING DETECTORS

The X-ray exposure to the patient in CT scanning has been of major concern for radiologists and physicists as the number of these examinations continues to increase. One study showed that, in the United States, 11% of diagnostic radiological procedures are CT examinations; however, their contributions to the collective patient dose from diagnostic radiology are as much as 67%.[5] Therefore, new methods that reduce the patient dose in CT examinations while keeping the noise and quality of images at the same level as conventional methods will have a significant impact on public health.

When tissues have similar linear attenuation coefficients—for example, fibrous tissue and a tumor, or bone and iodine-enhanced blood—it is difficult to distinguish between them (or to identify the tissue type) using conventional X-ray CT scanners.[6] Contrast resolution can be improved by increasing the signal-to-noise ratio (SNR), which can be improved by increasing the X-ray tube current. However, this is not appropriate in a clinical setting in which radiation dose to patients is already a concern with this conventional CT methodology. Thus, one of our goals is to develop a method to reduce noise and therefore improve the SNR in CT images by using photon-counting detectors that have a lower noise floor than current-mode CT detectors.

There are several clinical CT applications where patients would benefit from a lower dose CT scan that maintains sufficient contrast to obtain an accurate diagnosis. For example, lung cancer is the second leading cause of death in the United States following heart disease.[7] The American Cancer Society estimates that 22.7% of the deaths that will occur in 2006 will be from cancer and that roughly 30% of these cancers will begin in the lungs.[8] Lung cancer kills more people than colon, breast, and prostate cancers combined. In the United States there are an estimated 160,000 deaths from lung cancer, with 172,000 new cases detected each year.[9–10] For these reasons, CT has been investigated for lung cancer screening.[11] Recent studies indicate that CT has an increased capability to detect small malignant nodules as compared to chest X-ray imaging.[12–13] However, radiation dose to the patient is still a principal concern in the widespread use of CT for screening.[14,15,16,17,18]

Another application that would benefit from a lower dose CT is colon cancer examinations. Colorectal carcinoma is the second most common cause of cancer mortality in the United States. It is estimated that this cancer is associated with 60,000 deaths annually, with 150,000 new diagnoses every year.[19] Most nonhereditary colorectal carcinomas originate from preexisting adenomatous polyps. If early detection and

removal of these polyps can be performed, it will significantly reduce the morbidity and mortality.[20] Clinical studies have shown that the sensitivity and specificity of CT colonography averaged 75% and 90% in patients with adenomas 10 mm in diameter or larger.[21] The risk of malignancy is 10% for adenomas 10–20 mm in diameter and increases at least 30% in adenomas larger than 20 mm in diameter. Because CT colonography offers a complete examination of colon in a very short time, is noninvasive, and can be used without sedation, it may become the screening examination of choice due to its high sensitivity and specificity. Photon counting with optimal energy weighting can be applied to increase soft-tissue contrast at reduced dose for CT colonography. In addition to soft-tissue contrast-resolution enhancement with optical energy weighting, high-contrast resolution can be increased by subtracting low- and high-energy images.[22]

To improve contrast at reduced dose, photon-counting detectors have been considered as alternatives to conventional X-ray detectors used in radiography and CT.[23,24] Photon-counting X-ray detectors applied to mammography and radiography have shown a dose reduction of 40% to 400% while maintaining sufficient contrast for these applications.[25] Simulations have shown that photon counting along with optimal energy weighting can increase dose efficiency up to 40% for digital mammography as compared to a conventional integrating (current-mode) detector system.[26] One group has reported that the lesion-detection SNR is significantly affected by the energy dependence of the detector's quantum efficiency and the varying contrast carried by different energies in the beam spectrum.[27] Recent studies have demonstrated that weighting photons with $E^{-3}$ can lead to an improved SNR compared to simple photon counting or linear energy weighting (as in intensity-integrating detectors). In another simulation study, it was demonstrated that a photon-counting detector can reduce beam hardening and improve lesion SNR ratio compared to conventional detectors; optimally weighting the photons further improved the SNR but resulted in greater beam hardening due to the greater weight given to lower energy photons.[28] A photon-counting detector with energy binning could also potentially benefit dual-energy CT studies for imaging iodine and calcifications in arteries, to improve CT angiography,[29] and enhance bone densitometry.[30]

## 3.4 ENERGY INFORMATION IN CT WITH PHOTON-COUNTING DETECTORS

The use of photon energy in CT images has the potential to provide additional information, as attenuation coefficients of materials vary differently with the applied energy (see Figure 3.1). The basic idea of multiple-energy imaging is to use this variation to analyze and identify materials in the scanned object.[31] Energy information has been applied to mammography for enhanced visualization of microcalcifications, chest radiography to enhance soft-tissue contrast,[32,33] bone-mineral densitometry to provide accurate measurement of bone-mineral density,[34] and enhanced contrast-agent visualization.[35] For example, two images can be created, one for soft tissue and the other for bone. Alternatively, one can extract parameters of physical interest, such as the fraction of attenuation in a pixel that

is due to photoelectric or Compton interactions. Using these compositions, one can create attenuation map images for any desired incident X-ray energy. This is useful for attenuation compensation in positron emission tomography (PET) and single-photon-emission computed tomography (SPECT), or as an input to radiation dosimetry treatment plans. In addition, these images can be useful to provide information about small features such as microcalcifications, calcified plaques in arteries, and bone-mineral content.[36]

The combination of these two advantages obtained with photon-counting detectors, namely the reduction of beam-hardening artifacts and the identification of materials, will be significant for many clinical applications. Most vascular imaging involves contrast agents such as iodine or gadolinium. Patients with vascular disease often develop calcium deposits (hard plaque) adjacent to vulnerable, fatty soft plaque. Because such vulnerable plaque may rupture and cause a stroke, it is crucial to detect and characterize. Detecting and characterizing plaque is difficult with current-mode detectors. First, calcium and contrast-enhanced blood are difficult to distinguish with current-mode detectors because they have similar attenuation coefficients. Second, calcium deposits may generate a strong beam-hardening artifact that looks similar to soft plaque, making characterization difficult. Therefore, the advantages of photon-counting detectors over current-mode detectors can substantially improve the accuracy of diagnosis and treatment planning in vascular imaging (e.g., carotid artery imaging).

There are two basic types of information that can be extracted from energy-resolved CT images, namely physical and material. Methods that extract physical information decompose images into components that have a physical meaning. These can be either the photoelectric and Compton components[37] or the electron density[38] and effective atomic number, $Z_{\text{eff}}$.[39] One limitation of these techniques is that they require assumptions about the underlying physical phenomena. For example, the photoelectric component is assumed to vary as $Z^s E^{-r}$, where $Z$ is the effective atomic number, $E$ is the X-ray energy, and $s$ and $r$ are empirical constants. However, it has been found that both $s$ and $r$ depend on energy and $Z$. Moreover, this assumption is not valid at the K-edge, where the change of the photoelectric component is discontinuous.

An alternative method that extracts material information decomposes images into basis materials.[40] The materials are typically chosen based on a combination of factors, including convenience in performing calibrations and how closely they match materials of interest in the object. For medical imaging research with phantoms, common choices of materials include polymethyl methacrylate (PMMA) for simulating soft tissue, and Al for bone. Note that after obtaining a basis decomposition of the object, it is possible to use this information to obtain either electron density and effective $Z$ images or basis decompositions in terms of other pairs of materials.

In an attempt to take advantage of the benefit of multiple-energy CT images, several dual-energy CT systems (sometimes called *dual kVp CT systems*) have been developed using conventional current-mode detectors.[41] To date, several methods have been employed to obtain dual-energy images with current-mode detectors. One method involves taking multiple scans with different attenuating filters placed in the X-ray beam before it reaches the patient. Another method involves using a

set of two X-ray tubes either with two arcs of detectors or a two-layered detector sandwiched together, which can register both a lower X-ray energy band in the layer closer to the patient and a higher energy band (transmitted through the first layer) in the second layer. Another approach utilizes a fast switching of the voltage applied to the X-ray tube.

All these methods suffer from one or more problems. These problems include the following:

1. A misregistration between multiple data sets obtained with different energies in the presence of patient/organ motion will degrade the spatial resolution of the fused images.
2. Co-scatter exists between the two detector rings or layers, and these systems are expensive due to the additional X-ray tubes and detectors required.
3. Switching to a lower voltage applied to the X-ray tube causes a significant reduction in the X-ray flux, leading to more statistical errors.
4. An additional limitation inherent in all these methods is that the multiple-energy data sets are still energy weighted within each data set.

Photon-counting detectors can ameliorate these problems. In addition, applying a uniform weight to each photon with a single photon-counting detector is preferable for noise reduction, instead of applying a weight proportional to the energy deposited as is done with current-mode (intensity integrating) types of detectors. This is important because it has been shown that one type of object composition analysis method required a noise level as low as 1/3 to 1/10 of the noise encountered in standard CT imaging.[42]

## 3.5 PHOTON-COUNTING X-RAY DETECTORS IN MEDICAL IMAGING

Photon-counting detectors with energy-discrimination capabilities have been developed for nuclear medicine applications such as PET and SPECT and also for X-ray diffraction imaging. They also have the potential to address both of the two major problems inherent in current-mode CT imaging, namely, dose and contrast. However, in CT there are very demanding requirements on the detector for particularly high count rates as well as good detection efficiency and reasonable energy resolution. To meet these requirements, there is a need for the development of novel detector structures, customized fast low-noise and low-power ASIC electronics, and appropriate interconnections between detector pixels and electronics for the construction of tileable detector modules.

In an effort to overcome the limits imposed by current-mode detectors, a number of research groups in hospitals, universities, and commercial companies have already started to develop photon-counting systems for X-ray imaging applications.[43] For example, Philips Research Laboratories has recently presented preliminary results from an X-ray detector that counts the X-ray flux.[44] A group at the University of Erlangen-Nurnberg recently reported on their efforts to use the photon-

counting Medipix2 detector to perform material decomposition in X-ray imaging. GE Healthcare has populated a GE CT scanner with prototype photon-counting detector modules designed and fabricated by the authors at DxRay, Inc. The prototype photon-counting detector modules have generated the first clinical images using a fully photon-counting CT system and are described in detail in Section 3.12.

In a mammography study, Si strip detectors that were wire bonded to photon-counting electronics were incorporated into a scanning slit geometry.[48] Because of the small stopping power of the thin Si, the detectors had to be oriented so that the X-ray beam passes through them edge-on. The spacing between the readout lines on the Si was 50 μm. The X-ray source, collimators, and detector readouts were all mounted on a common mechanical structure that performed a scanning motion to create an image. A dose reduction up to 40% can be obtained by making use of photon counting and energy weighting, as compared to a conventional mammography system. For example, computer simulations have shown[45] that photon counting combined with an optimal energy weighting scheme[46,47] increases the dose efficiency up to 40% compared with a signal-integrating system.[48] Photon counting alone was shown to account for half of this increased performance. The development of all of these photon-counting X-ray imaging systems is made possible by advances in interconnecting technologies to couple custom ASIC readout systems to pixelated semiconductor detectors.

## 3.6 DIRECT-CONVERSION X-RAY DETECTORS

As mentioned previously, conventional CT systems typically use a detector comprised of a pixelated Si photodiode that is optically coupled to a segmented scintillation crystal. In these detectors, the energy of an absorbed X-ray is converted to an electrical charge by an indirect method, first generating light photons in the scintillator, and then in turn the light photons are converted into an electrical signal in the photodiode (see Figure 3.2). In contrast, a direct-conversion detector makes use of the direct conversion of the X-ray photon energy deposited by each X-ray to charge. Counting and sorting X-ray photons with different energies is accomplished by the associated amplification and processing electronics for each pixel (see Figure 3.3). Direct conversion with photon counting offers numerous advantages applicable in various X-ray imaging applications.

Single crystals of CdTe and CZT can effectively absorb the incoming X-ray radiation and convert it directly into electrical signals, which can be read by an ASIC. Thus the overall signal-to-noise ratio can be significantly better than that achievable with indirect detectors utilizing photodiodes. The efficiency of converting the X-ray signal to an electrical signal in CdTe and CdZnTe can be much higher than that in the scintillation detectors due to the basic underlying physics of the energy-transfer process in the direct-detection approach. That is, the mean energy for creating an electron-hole pair in a semiconductor detector (4.43 electron-Volts per electron-hole pair in CdTe) is typically much smaller than the corresponding energy necessary to create an optical photon and consequently an electron-hole pair in the photodiode through the scintillation approach. Another important consideration is that charges generated by X-rays do not spread laterally (aside from negligible diffusion), but

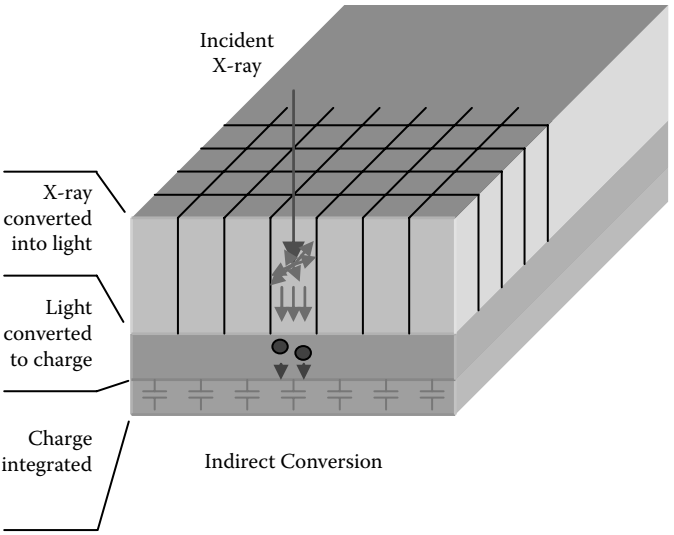



**FIGURE 3.2** Schematic of indirect-conversion X-ray detector. Each X-ray generates visible-light photons, some of which are converted to electric charge. The charge is integrated by readout electronics.

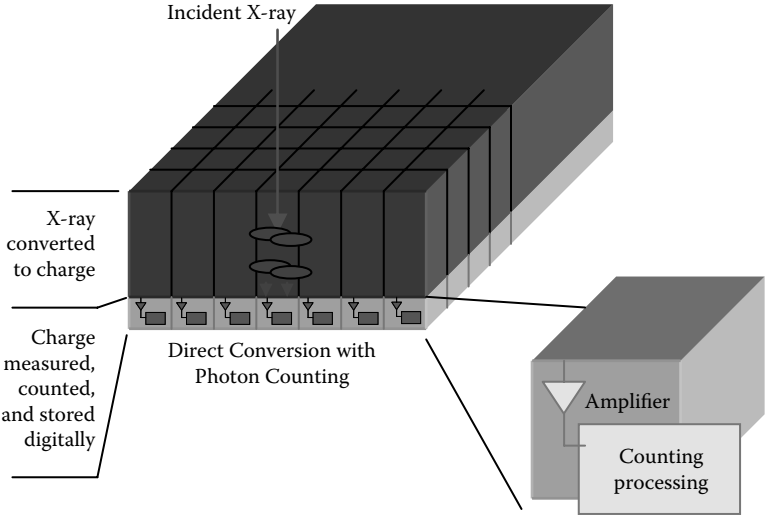



**FIGURE 3.3** Schematic of photon-counting direct-conversion X-ray detector. Each X-ray generates an electric pulse. The pulse is processed, and qualifying pulses increment a counter at the pixel.

move along the applied electric field lines within the semiconductor. Spreading of light in the indirect conversion method is limited by segmenting the scintillator. However, for small segments less than or equal to 1 $mm^2$, the physical segmentation and surface treatment for the required light reflection contribute to appreciable loss in the light output of the scintillator.

## 3.7 CADMIUM TELLURIDE AND CADMIUM ZINC TELLURIDE DETECTORS

Recent advances in interconnecting technologies to couple custom ASIC readout systems to pixelated semiconductor detectors are used to create direct-conversion X-ray imaging detector arrays. This method requires pixelation. There exist several high-atomic-number semiconductor-detector materials capable of pixelation, such as high-purity germanium (HPGe), cadmium telluride (CdTe), and cadmium zinc telluride (CZT), that are capable of being bump-bonded to a ball grid array (BGA) containing ASICs. HPGe requires cryogenic cooling to reduce thermally generated dark current, which is a dominant source of noise in these detectors. CdTe and CZT, on the other hand, do not require any cryogenic cooling. For X-ray imaging applications, the significant properties of these materials are that they have high stopping power for X-rays, making the detectors very efficient, and they have low leakage current at room temperature, making intrinsically low-noise devices. CdTe and CZT are very efficient direct X-ray converters, yielding a very large signal.

Another very important parameter is the ability to quickly and efficiently collect charges liberated by the incoming radiation. Mobility-lifetime ($\mu\tau$) products for electrons and holes are measures of these abilities. Despite the utilization of special detector structures that make use of the "small-pixel effect" or drift structures that favor electron-only collection for creation of an output signal, the collection of holes is also important, as discussed below. By making use of these special detector structures, a response signal can be much faster than the transit time of an electron (or hole) over the whole detector thickness, thus allowing for very fast counting and thus the ability to cope with high X-ray flux.

The extremely high X-ray photon-count rates required by CT impose a number of requirements on a photon-counting energy-dispersive CT detector, such as the design of the processing electronics, the selection of detector material, and the method of electrode fabrication. To avoid a space-charge buildup in CdTe or CZT crystals that would lead to distortion and eventual collapse of the electric field, care must be taken that the charge generated by the radiation is removed from the device at a sufficiently fast rate.[49] For this reason, hole collection (holes being less mobile than electrons) is critically important. In this respect, CdTe has some advantages over CdZnTe, as typically the lifetimes of holes are significantly longer in CdTe than in CdZnTe (see Table 3.1). CT applications require very high detector count-rate capabilities. These high count rates can be achieved by using very short shaping times in the amplification circuitry. Despite the slightly higher dark current in CdTe detectors, due to a lower resistivity of the material compared to that of CdZnTe, CdTe detectors do not have a disadvantage here because, at short shaping times, the electronic noise is dominated by series noise, and the parallel noise due to dark current can be neglected.

The potential to increase signal with direct conversion has led to the investigation of several other high-Z semiconductor materials for X-ray imaging, such as thallium bromide (TlBr),[50] lead iodide ($PbI_2$),[51,52] and mercuric iodide ($HgI_2$).[53,54] Table 3.1 shows the electrical properties of these semiconductor X-ray converters. The mean energy for creation of an electron-hole pair is a very important material parameter

**TABLE 3.1**
**Characteristics of Candidate Detector Materials for X-ray Imaging: Analysis of Single Crystals at Room Temperature**

| Material | Mean Energy for e-h Creation (eV) | Density ($g/cm^3$) | Atomic Number | Resistivity (Ohm-cm) | Electron Mobility-Lifetime Product ($cm^2/V$) | Hole Mobility-Lifetime Product ($cm^2/V$) |
|---|---|---|---|---|---|---|
| CdTe | 4.43 | 6.2 | 48,52 | $10^9$ | $3.3 \times 10^{-3}$ | $2 \times 10^{-4}$ |
| CdZnTe | 5 | 6 | 48,30,52 | $10^{11}$ | $1 \times 10^{-3}$ | $6 \times 10^{-6}$ |
| HgI2 | 4.2 | 6.4 | 80,53 | $10^{13}$ | $10^{-4}$ | $4 \times 10^{-5}$ |
| TlBr | 6.5 | 7.56 | 81,35 | $10^{12}$ | $1.6 \times 10^{-5}$ | $1.5 \times 10^{-6}$ |
| PbI2 | 4.9 | 6.2 | 82,53 | $10^{12}$ | $8 \times 10^{-6}$ | |

that is related to the efficiency in energy transfer of X-rays into ionized charges. A material with small conversion energy will produce a larger number of ionized charges, and therefore a signal with improved statistical characteristics. The $\mu\tau$ product is a measure of the charge-collection efficiency in the material, and larger values of this parameter ensure that more of the ionized charges are collected and larger resulting signals can be obtained. The last column in Table 3.1 lists the $\mu\tau$ product for holes. Additional important parameters characterizing these materials are not shown in the table, including the cost and degree of difficulty involved with the production of good quality and stable material. TlBr, $HgI_2$, and $PbI_2$ are soft and have mechanical and physical properties making them very difficult to use with bump-bonding technologies. Based on these considerations, CdTe and CdZnTe clearly are the good candidates for X-ray CT imaging applications in view of the overall values of parameters listed in Table 3.1.

## 3.8 PHOTON-COUNTING CADMIUM TELLURIDE DETECTOR DESIGN AND FABRICATION FOR CT

To develop a clinical photon-counting energy-dispersive CT detector technology, we have designed and fabricated vertically integrated direct-conversion pixelated semiconductor X-ray imaging detector modules. The modules are formed by bump-bonding pixelated CdTe crystals to a BGA package containing readout ASIC electronics. The vertical integration allows tiling in two dimensions to create any number of pixel rows in a fan-beam multislice CT scanner.

The CdTe detectors are processed using specific surface preparation and lithography techniques. One important factor that affects the performance of CdTe detectors is the interface between the CdTe bulk material and the metal contact. The cutting and dicing process inevitably leaves a high density of electrically active defects in the near-surface region of the crystal.[55] Typically, a solution of bromine–methanol is used to etch the surface of the CdTe subsequent to the dicing process.[56] However, bromine–methanol appears to leave a Te-rich surface, which then rapidly oxidizes

after exposure to air.[57,58] The thin oxide layer affects both the bulk and surface leakage-current characteristics, and the interface properties vary considerably from process to process. It is preferable to obtain a stoichiometric surface that has more reliable surface-leakage characteristics. This can be done by mechanical polishing of the surface and avoiding the bromine-methanol etch. After surface polishing is complete, the samples are kept in a dry $N_2$ or vacuum environment to stabilize the surface. The metal contacts can be formed using Pt or Au. The metal is patterned using a "lift-off" technique, in which the photoresist is first spun onto the CdTe surface; then the resist is exposed using the desired pattern in a glass photomask. The resist is developed and removed in areas where the metal will contact the CdTe surface. After the metal is deposited, the remaining resist is then wet-etched away, which also removes the unwanted metal on top of the resist.

To optimize the thickness of a CdTe detector for CT, the detection efficiency for diagnostic X-rays should be maximized within a total charge-transport time close to the shaping time used. Calculations of the percent attenuation were performed as a function of energy for CdTe at several thicknesses between 1 and 5 mm. Clinical CT scans typically use between 120 and 140 kVp for the X-ray source, and the polyenergetic spectrum produced then has a maximum output of X-rays between approximately 70 and 80 keV, respectively. Taking into account electron transit time in the bulk of the detector, the thickness was selected to be 3 mm and the pixel pitch was 1 mm. This way, the electron transit time can be limited to about 60 ns (at −1000-V bias), and the resulting signal rise time (fast portion) induced by the small-pixel effect[57] at the readout electronics will be about 20 ns. This value of the signal rise time will be useful with 30-ns shaping times and at the same time maintain approximately 99% absorption at 80 keV.

With surface preparation, thin-film deposition, and detector-thickness methods designed for production of a fast photon-counting X-ray CdTe detector module, attention must be turned to the exposed lateral surfaces of the CdTe crystal to ensure reliable operation at high-bias voltage without large surface-leakage currents. Also, the noise performance of individual pixels in single-crystal multi-anode (pixelated) CdTe detectors is limited, in part, by the contribution of current present when the detector is biased to a high voltage and no ionizing radiation is present. This dark current usually is larger, per unit pixel area, for edge pixels, and usually the edge pixels are responsible for premature detector breakdown. To minimize surface-current contribution, a ring of conductive material, a guard ring, is often placed around the pixels close to the edge of the crystal and connected to ground. However, guard rings reduce the effective area of the detector and introduce an additional dead space when many detectors are tiled together.

To circumvent this problem, we have developed a method of passivating the lateral surfaces of CdTe detectors and eliminated the need for guard rings.[60] Reducing surface current on these crystals without the use of guard rings could allow pixel maps to be extended to the edge of the crystal, thus making more effective use of the material and allowing tiling with little or no dead space. Figure 3.4 shows a 3-mm-thick CdTe crystal configured with a 256-pixel anode at a 1-mm pitch, where surface preparation and film deposition were performed as described previously. The perim-

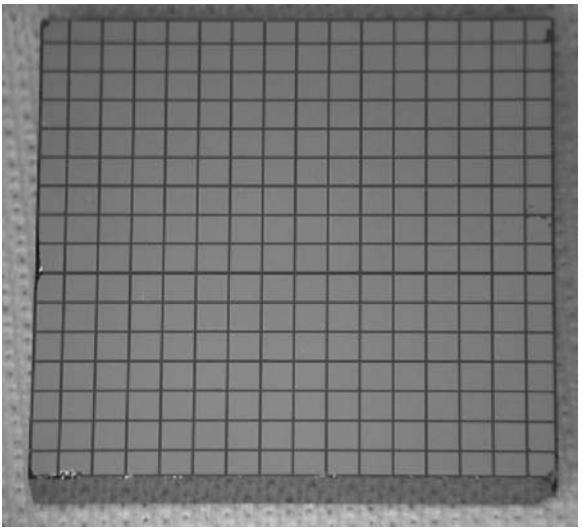


**FIGURE 3.4** An image of a 256-pixel CdTe detector used for clinical CT.

eter pixels are slightly smaller than the interior pixels, so that when several detectors are tiled together, a 1-mm pixel pitch can be maintained.

## 3.9 PHOTON-COUNTING CdTe DETECTOR CHARACTERISTICS

To test the CdTe pixel detectors for their electrical performance, a device with 256 spring-loaded electrical contacts connected to measurement instruments was used. Figure 3.5 shows a schematic of the 256-pin testing device that we have made that allows us to test each pixel with discrete electronics of our choice.[61] For a per-pixel dark-current measurement, a detector is placed in the custom jig with the anode (pixel) surface facing down. Such placement allows the illumination of detectors by X-rays or radioactive sources. This jig can ensure repeated reliable electrical connections with the pixels without mechanical damage to the surface metallization.

Pixel selection is performed by a selector board, which consists of 78 single-pole, double-throw (SPDT) switches controlled by a personal computer over a USB port. With this setup, measurements of the leakage current versus bias voltage (I-V characteristic) can be performed for each pixel of the detector. In addition to measuring per-pixel dark-current values, the selector board can connect each pixel to external electronics, allowing the implementation of photon counting with the same design as an ASIC.

To perform photon-counting measurements, the test setup uses commercial fast low-noise amplification electronics and a multichannel analyzer to measure spectra from radiation sources. Radiation sources can be used in signal rise-time measurements. The monoenergetic gamma rays produced by radionuclide sources allow the measurement of the spectroscopic capability of the detectors. The energy resolution is defined as a full width at half of maximum (FWHM) of the measured pulse height distribution of the pulses. The energy resolution is measured as a function of the

shaping time (filter) of the amplifier and as a function of the radiation energy. The fast filter amplifier ORTEC Model 579 allows varying shaping times from 10 to 500 ns. The detectors produced are tested in realistic conditions by being exposed to X-ray flux from an X-ray generator designed for CT.

### 3.9.1 CdTe Detector Dark-Current Measurements

To develop our passivation method and to characterize the individual pixels of the CdTe crystals, many dark-current measurements have been performed in the testing apparatus shown in Figure 3.5.

Figure 3.6 shows the individual dark currents of a 256-pixel (16 × 16) CdTe detector before passivation. The graph shows the pixels in sequence one row at a time; therefore, the first and last 16 pixels in the plot are on the edge, and the other edge pixels appear in intervals of 16. The edge pixels are clearly identifiable in Figure 3.6, as they have not been passivated and therefore have higher dark current than the central pixels. The four highest dark-current values in Figure 3.6 are the corner pixels,

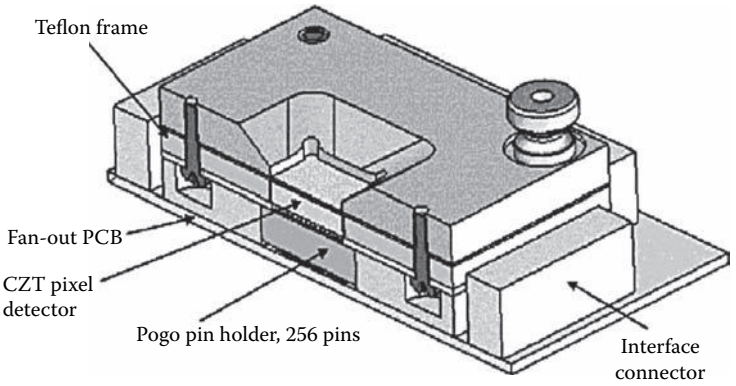



**FIGURE 3.5** A drawing of a 256-pogo-pin test jig to supply electrical contact to each pixel in the detector array.

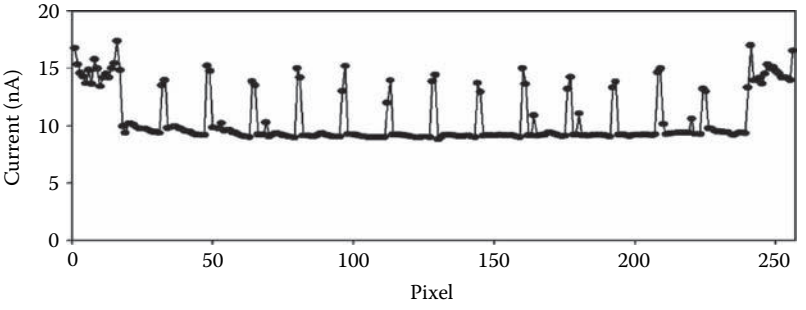



**FIGURE 3.6** Dark current versus pixel number for a 256-pixel CdTe detector without any passivation of the lateral sides.

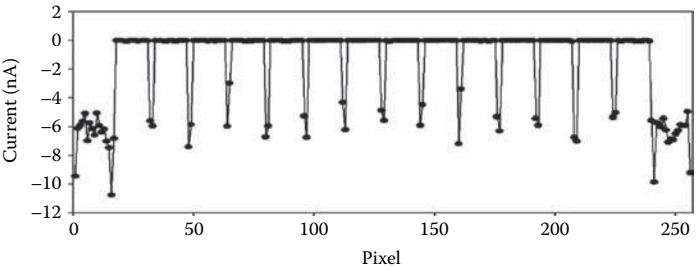



**FIGURE 3.7** When the detector in Figure 3.6 is passivated, a significant amount of surface current is eliminated. This graph shows the difference in dark current before and after passivation versus pixel number for a 256-pixel CdTe detector with passivation of the lateral sides.

which each have two lateral surfaces. These corner pixels can be found at 1, 16, 241, and 256 along the "pixel" axis in the plot.

Figure 3.7 shows the difference in dark current before and after passivation and indicates a dramatic reduction in dark current for the edge pixels without affecting the central pixels. The reduction in dark current is largest for the four corner pixels. This is to be expected, since the lateral surface area is largest for these. A comparison of dark current between edge and central pixels after passivation demonstrates that the value for the edge pixels is proportional to the pixel area with the same proportionality constant as for the central pixels. This indicates a nearly complete suppression of surface current after passivation. This method is applied to the CdTe detectors and allows for the elimination of the need for guard rings.

### 3.9.2 Fast Photon Counting with CdTe and Discrete Electronics

To characterize charge collection properties, individual pixels have been connected to discrete electronics using the apparatus in Figure 3.5. Figure 3.8 presents a typical pulse shape from a detector pixel obtained in response to a gamma-ray photon event from a $^{57}$Co source. Note that the fast portion of the pulse has about a 20-ns

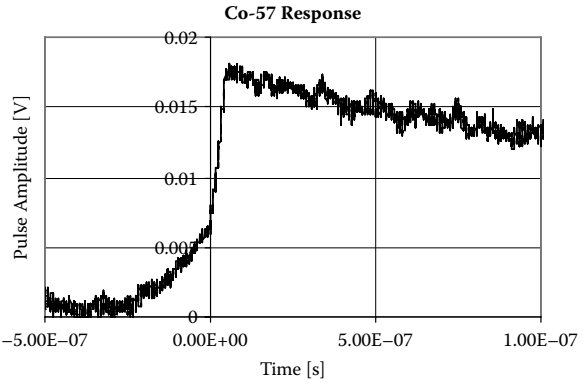



**FIGURE 3.8** A pulse shape measured for a detector pixel. (© 2009 by IEEE. With permission.)

rise time. Figure 3.9 shows the spectral response of one small pixel to gamma-ray photons from a $^{57}$Co source taken with a customized commercial discrete electronics amplifier and a multichannel analyzer. Despite a very short peaking time of 30 ns, an energy resolution of 4.75% (5.8 keV FWHM at 122 keV) is obtained. The electronic noise is 4.8 keV (FWHM), as measured with an electronic pulser. We have chosen a 30-ns peaking time in the amplifier to accommodate both the rise-time signal and high flux rates. In order to test this we have employed a delay line shaper fed into an ORTEC 579 fast filter amplifier. Figure 3.10 contains a screen capture

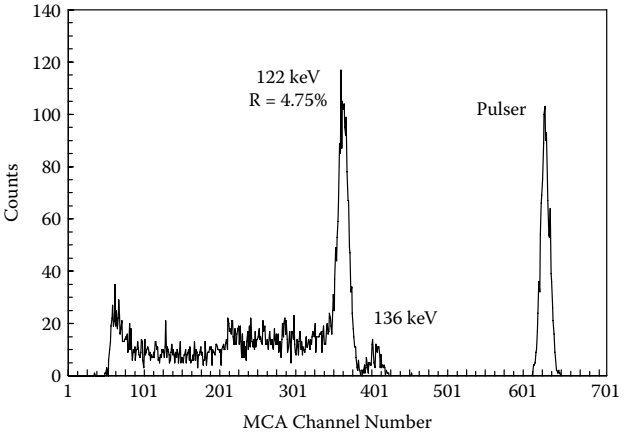



**FIGURE 3.9** Spectrum taken with a single detector pixel in response to Co-57 source using a short shaping time. Electronic noise is represented by pulses injected from an external generator (pulser peak). (© 2009 by IEEE. With permission.)

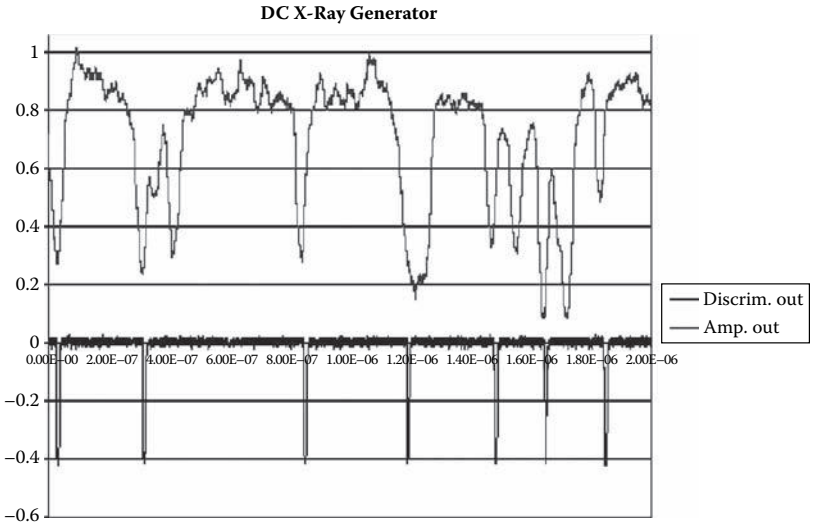



**FIGURE 3.10** A screen capture of a digital oscilloscope showing the output of the shaping amplifier (at top) and the output of the discriminator (at bottom).

of an oscilloscope that shows the response of the amplifier and discriminator stage as a function of time with exposure to X-rays from a clinical X-ray tube operated at 140 kVp. The upper trace in Figure 3.10 shows the negative pulses at the shaping amplifier output, and the lower trace shows the output of the discriminator.

These measurements have been performed at a very high flux of up to $6 \times 10^6$ output counts per second per pixel. Pulse pileup or dead-time losses eventually limit the upper bound on the flux linearity of the detector and discrete readout. Testing with discrete electronics verifies the rapid charge collection required for fast energy-dispersive photon counting and that the design of our CdTe detectors is suitable for bonding to a multichannel ASIC readout. In designing the ASIC, vertical integration of the readout with the active area of the pixels is required for tiling in two dimensions. Two-dimensional tiling allows modules to form multirow fan beam detector arcs of any size.

## 3.10 PHOTON-COUNTING ASIC ARCHITECTURE FOR CT

The implementation of a fast photon-counting ASIC is significant for several reasons. By combining the amplification, pulse shaping, and readout for multiple electronic channels into a single chip placed behind the detector, the CdTe crystals can be tiled with very little dead space, allowing for construction of practical imaging systems. Our ASIC contains 128 parallel readout channels, each connected to one CdTe pixel. All the channels are equipped with an amplifier, a shaper, two discriminators, two 5-bit digital-to-analog (DAC) converters, and two counters. Figure 3.11 shows the basic architecture of an individual channel.

Each channel uses a transimpedance amplifier with an active feedback. The input stage is based on a classical cascode configuration with an nMOS (n-type metal oxide semiconductor) input transistor and a pMOS (p-type metal oxide semiconductor) current source load (see Figure 3.12). The input transistor size is optimized to match 5 pF of combined capacitance from the detector plus the parasitic part. For minimum noise, the nominal bias of the input transistor is 500 μA. The amplifier is AC-coupled to a shaper, which uses a voltage amplifier in a single-ended configuration of two cascaded common-source stages. It provides additional amplification and introduces the dominant integration time constant. The gain at the shaper output is

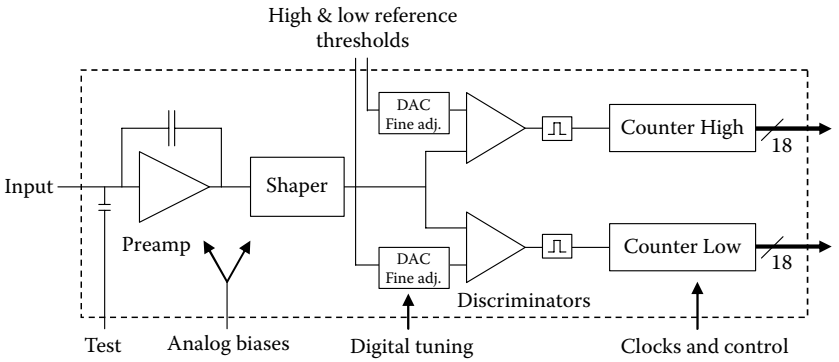



**FIGURE 3.11** Channel architecture. (© 2009 by IEEE. With permission.)

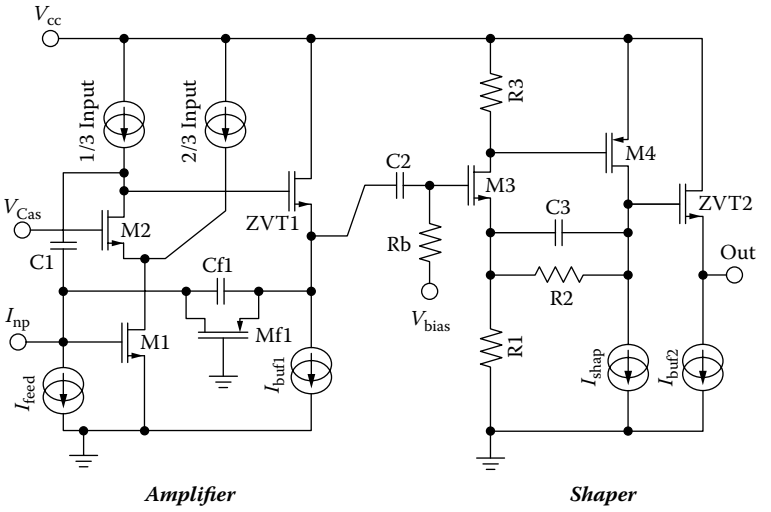



**FIGURE 3.12** Principle schema of the amplifier. (© 2009 by IEEE. With permission.)

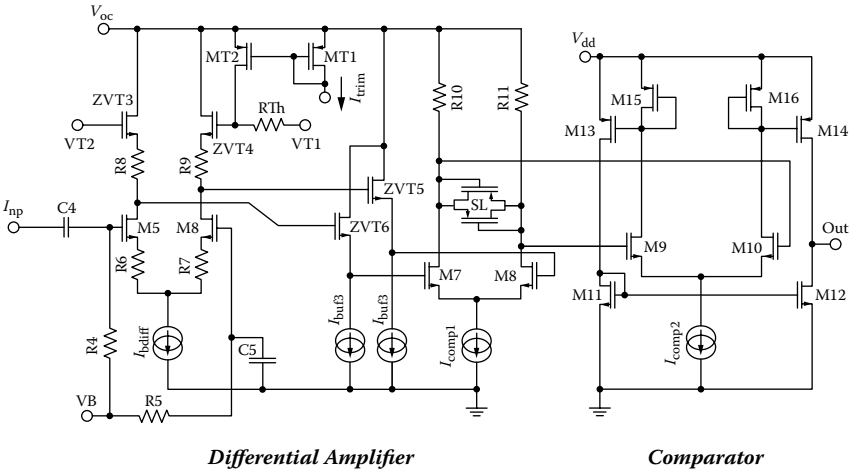



**FIGURE 3.13** Principle schema of discriminator circuit. (© 2009 by IEEE. With permission.)

about 120 mV/fC, and the signal gives a peaking time response on a delta pulse at the input of about 20 ns.

The shaper is followed by two discriminators with identical circuits. Figure 3.13 shows the principal schematic of one discriminator circuit.[62,63] The threshold is applied differentially, producing a DC shift at the output of the differential pair. The first threshold line (VT1) is connected to a 5-bit DAC, while the other (VT2) is common to all the channels. The DAC allows for fine tuning of the thresholds per channel to compensate for channel-to-channel variations caused by processing parameter fluctuations. The DACs are linked in a series daisy chain, and the values

are programmed by a series shift register. The differential signals are then buffered by two nMOS source followers to the comparator stage, which consists of a classical two-stage amplifier with very high DC gain. Each discriminator circuit is followed by an 18-bit counter designed in the form of an asynchronous binary type using static master-slave D-flip-flops in each cell.

The counters are gated in parallel by an external signal, and the readout is controlled by a sequential bit register, retrieving the contents of each 18-bit counter fully in parallel. The readout sequencer can operate at 50 MHz, allowing for a complete readout of all the $2 \times 128$ counters in about 5 μs. The full chip architecture is illustrated in Figure 3.14.

The peaking time was measured as a function of the detector capacitance.[64] For detector capacitances up to 10 pF, and in response to a near-delta pulse at the input, the signal showed a peaking time of ≈20 ns. Figure 3.15 shows the equivalent noise

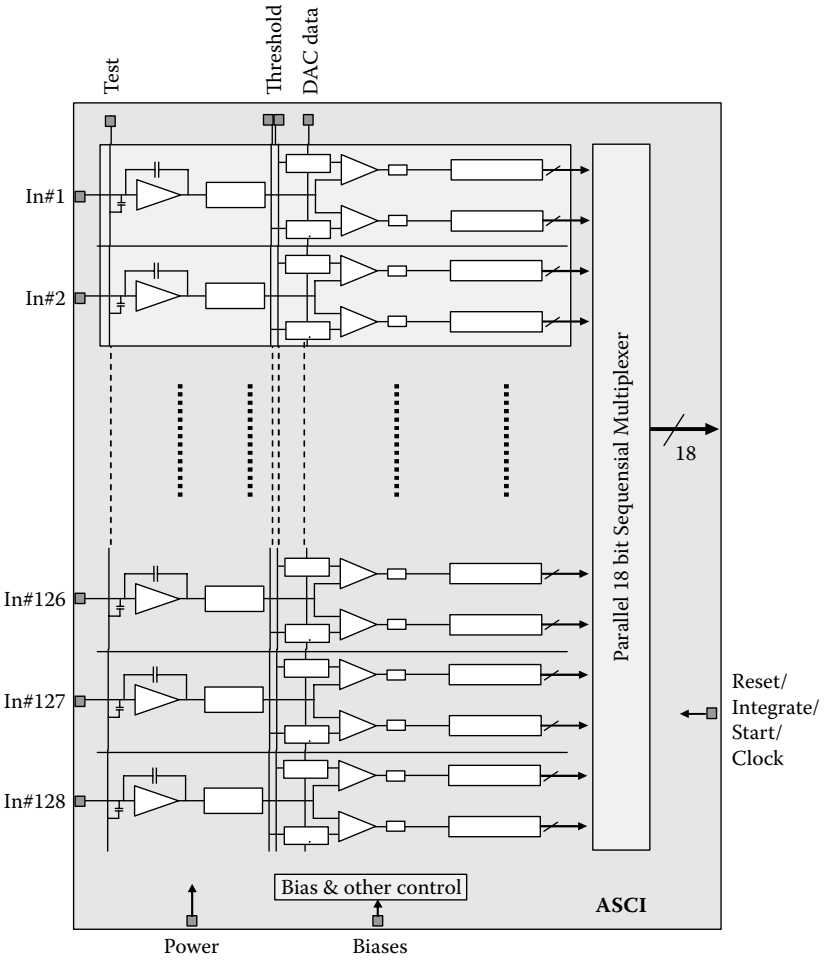



**FIGURE 3.14** The full ASIC architecture. (© 2009 by IEEE. With permission.)

charge (ENC) measured with the analog channel for different detector capacitances. At 0 pF, the noise is 430$e^-$, and the noise slope is measured to be 56.1$e^-$/pF. For the nominal detector capacitance of 5 pF, the noise was 711$e^-$. The measured value is higher than the calculated one, which can be explained by parasitic capacitances that were not taken into account.

The gain was measured by issuing a series of calibration pulses and scanning the threshold. All 128 channels were tested simultaneously at nominal bias currents. Figure 3.16 shows the gain distribution across the channels. The average measured

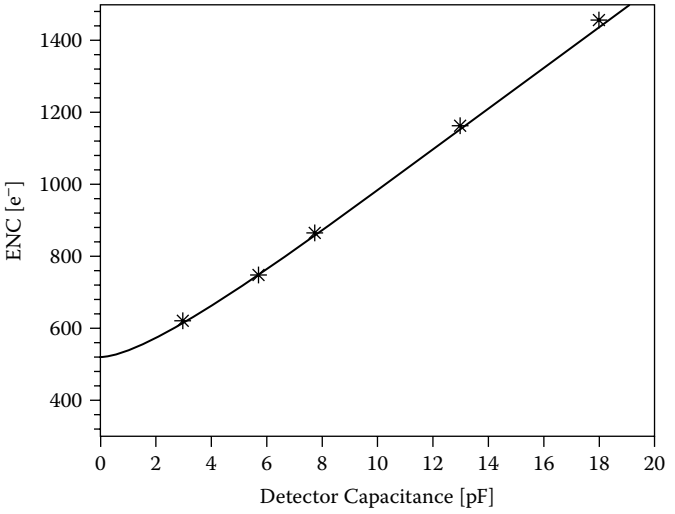



**FIGURE 3.15** Measured ENC [$e^-$ rms] versus input capacitance. (© 2009 by IEEE. With permission.)

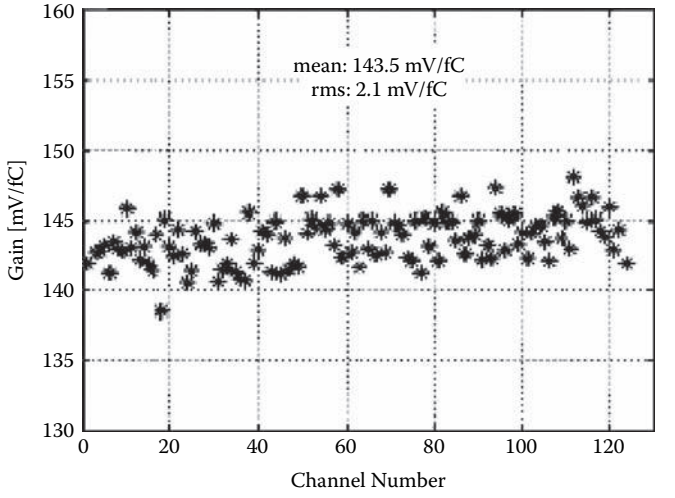



**FIGURE 3.16** Gain versus channel number. (© 2009 by IEEE. With permission.)

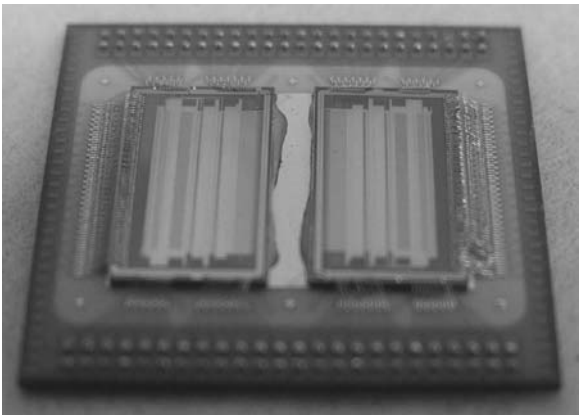


**FIGURE 3.17** A photograph of two ASIC chips (256 channels) assembled into a BGA package. The opposite side of the BGA is equipped with bonding pads to interconnect with the $16 \times 16$ CdTe detector pixels.

gain is 143.5 mV/fC, with a channel-to-channel gain variation on the order of 1.5% rms. The comparator offset was measured to 5.2 mV rms, which corresponds to 3.6% of the response equivalent to 1-fC input charge. When the chip was fully operating, a power consumption of 2.1 mW/channel was measured. Figure 3.17 shows a photograph of two ASIC chips (256 channels) assembled into a ball grid array (BGA) package. The silicon ASIC chips are wire-bonded to the BGA substrate. Two rows of bonding pads at the BGA serve as pathways for supplying power to the ASIC, providing control signals, and outputting the digital data. The opposite side of the BGA is equipped with bonding pads to interconnect with the $16 \times 16$ CdTe detector pixels.

## 3.11 PHOTON-COUNTING CT MODULE CHARACTERISTICS

First-generation photon-counting CT modules have been fabricated by bump-bonding CdTe detectors (see Figure 3.4) to fast photon-counting ASIC readouts (see Figure 3.17). The design of the CdTe detectors and the ASIC readout has been optimized for fast energy-dispersive X-ray imaging. The CT modules have been characterized for count rate performance and noise properties, and they have been used to produce clinical CT images.

## 3.12 PROTOTYPE CLINICAL PHOTON-COUNTING CT SYSTEM

To demonstrate the usefulness of photon-counting clinical CT, prototype detector modules were designed and fabricated with physical dimensions to specifically fit and populate a GE LightSpeed VCT scanner. The modules utilize the CdTe detector arrays, which make use of the small-pixel effect and are able to replace the conventional integrating detector units currently provided with the scanner. Each module included two CdTe detector arrays described in the previous sections to provide readout from 512 detector pixels, with a pixel grid of $1 \times 1$ mm.

Figure 3.18 shows a photograph of the fully assembled module. A top flexible tape provides bias voltage to the detectors. Both detector arrays are bump-bonded to the corresponding BGA packages containing ASIC electronics. In turn, BGA packages are interconnected with motherboard PCB (printed circuit board) and connector to receive power and provide I/O functions.

The modules were tested for electronic noise performance and count-rate capabilities. The low discriminator level (VT1) could safely be set above the noise level at 25 keV. The throughput characteristics were taken with a GE TTH Hawkeye X-ray generator as a function of the X-ray tube anode current. The X-ray generator was biased to 140 kVp, (peak kilovoltage) and X-rays were filtered through a 1-mm-thick Cu sheet. Figure 3.19 shows the output count rate (taken from the ASIC counter) of a typical detector pixel and its processing electronic channel with the low discriminator

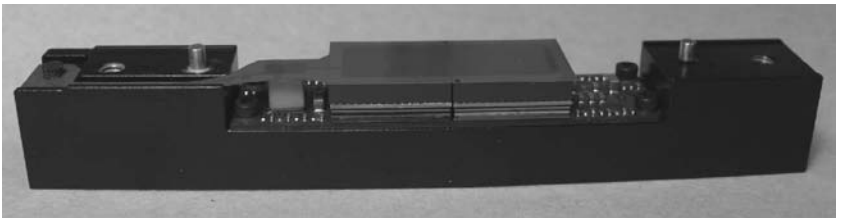


**FIGURE 3.18** A photograph of the fully assembled module used in GE LightSpeed VCT scanner.

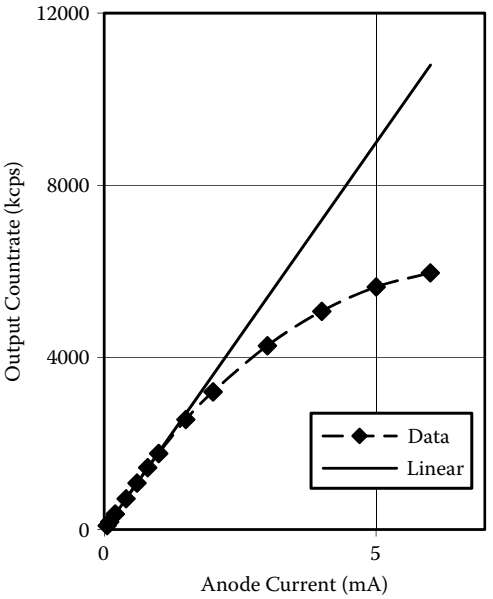



**FIGURE 3.19** Output count rate taken from a single detector pixel versus X-ray tube anode current. A straight line represents a linear fit to data taken at low count rates. (© 2009 by IEEE. With permission.)

level threshold set at 25 keV. The output count rate is fairly linear up to $3.5 \times 10^6$ cps (counts per second) and begins to saturate at about $6 \times 10^6$ cps. A linear fit to data taken at low count rates shows that the beginning of the saturation in the output count rate corresponds to about $11 \times 10^6$ cps at the input. By using corrections the output data can be made linear well above $3.5 \times 10^6$ cps, and by using this procedure, the useful dynamic range of the detector can be significantly expanded.[65,66]

Figure 3.20 shows an assembly of CT modules into a GE LightSpeed VCT scanner.

The conventional integrating detector modules of a clinical multislice scanner were replaced with the photon-counting modules shown in Figure 3.18. The detector system counts the X-ray photons transmitted through the patient and sorts the counts into two energy bins. The two bins of count data are processed to produce two sets of

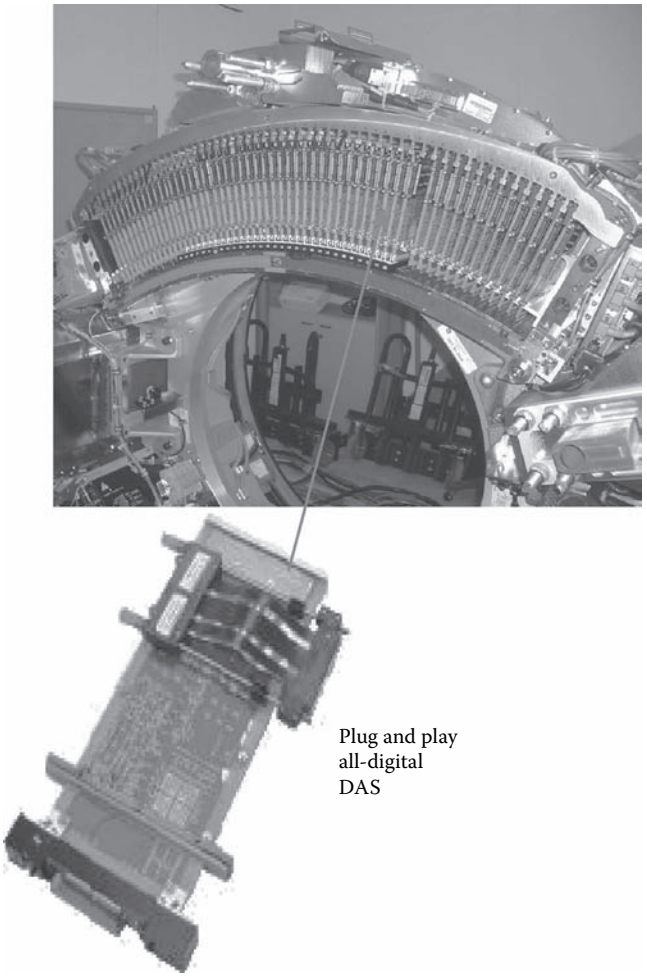



**FIGURE 3.20** Assembly of detector modules into GE LightSpeed VCT scanner.

raw data (high and low energy) that are input to the CT reconstruction algorithm. The raw data sets are processed with a dual energy calibration technique that enables projection space reconstruction and eliminates the beam-hardening effect. The image reconstruction generates mono-energetic equivalent CT images (with a user-selected energy) and "virtual-un-enhanced" images where the iodine-based contrast media is identified and removed in the reconstruction process (see Figures 3.21 and 3.22).[67] The images are presented in Hounsfield units (HU).

The first clinical trial with the prototype system was a prospective study on patients with known carotid artery disease. The study was performed at the Rabin Medical Center in Israel and was reviewed by the hospital IRB (Institutional Review Board) committee. The protocol was a CT angiography (CTA) scan of the carotid bifurcation region using a very low X-ray tube output (approx 1/10 the mA setting as compared to a conventional CTA exam). The axial images were processed to create MIP (maximum intensity projection), 3-D, volume rendering, and curve oblique images. Figure 3.23 illustrates the processed images. The clinical images demonstrate the vascular and stenotic elements with good quality, which is particularly notable in regard to the low X-ray tube current technique used for these scans.

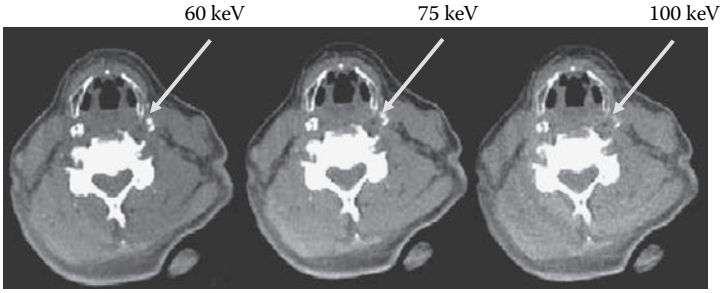



**FIGURE 3.21** Mono-energetic images reconstructed at 60, 75, and 100 keV (from left to right), respectively. The arrows show the left internal carotid artery. Note: the HU of the iodinated left internal carotid artery decreases with increasing energy. (© 2009 by IEEE. With permission.)

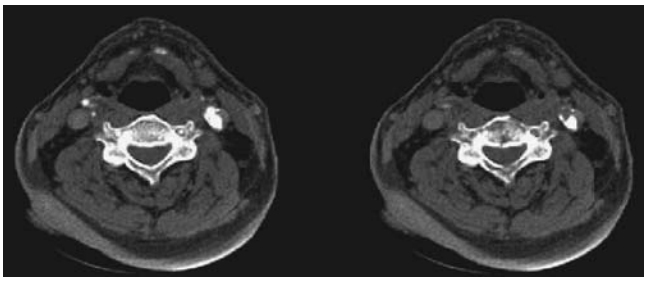


**FIGURE 3.22** Axial images at identical neck level. The left-side image shows contrast-enhanced neck arteries. The right-side image is the result of the unenhanced process (iodine is identified and removed). (© 2009 by IEEE. With permission.)

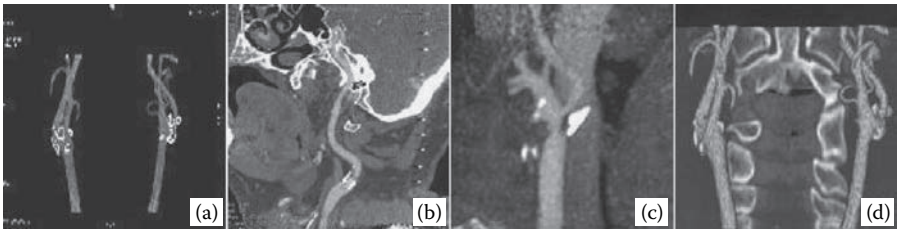



**FIGURE 3.23** Examples of the processed images: (a) 3-D, (b) curved oblique, (c) coronal MIP, (d) volume rendering. (© 2009 by IEEE. With permission.)

## 3.13 FUTURE WORK: DETECTOR ARRAYS WITH PARALLEL DRIFT STRUCTURES

X-ray throughput characteristics of photon-counting detectors are a limiting factor in X-ray flux (setting of current at the X-ray tube) that can be used without saturation of the system. This particularly limits the possibility of using such systems for imaging of moving organs (e.g., heart). Clearly, it is very important to develop detector structures with higher throughput characteristics. Our efforts to achieve higher throughput by applying parallel drift structures have shown the feasibility of this approach. Drift structures allow small anodes to collect X-ray-generated charge from an active area larger than the anode's area. We report our initial results are here. Before creating drift structures on CdTe crystals, we performed simulations of the electric field within the detector structure as well as simulations of the resulting output signal for different electrode patterns. We have been exploring the effect of changes to the following parameters: anode size, inner diameter of the contact that surrounds the anode, and the pixel size (thickness and width). Also, we varied the bias voltages applied to two electrodes (V1 = cathode potential and V2 = steering electrode potential). The anode potential is, by definition, kept at 0 V. It was assumed that the maximum voltage that we can apply to the detector cathode should not exceed 1000 V.

Based on these simulations, we chose to construct and test a detector array composed of 256 pixels (16 × 16) with an interpixel pitch of 1 mm and a detector thickness of 3 mm, where the anodes are formed as small 0.3-mm-diameter discs surrounded by a common steering electrode, and the cathode side is formed as an interconnected grid pattern.[68] Figure 3.24 shows photographs of the anode and cathode sides of a CZT detector array with this parallel drift structure.

Figure 3.25 presents a typical pulse shape from a single pixel of the drift-detector array obtained in response to a gamma-ray photon event from a Co-57 source. The steering electrode was biased to 200 V, whereas the cathode was kept at 1000 V. The pulse shape is shown in two time scales. The top and bottom drawings show the same pulse in 1000-ns and 10-ns scales per division, respectively. Note that the fast portion of the pulse is about 6 ns.

Figure 3.26 demonstrates the influence of different bias voltages applied to the steering electrode on the spectral properties of the detector. Spectra of a Co-57 radiation source were taken from a single pixel of the drift-detector array at bias

voltages of 0 V, 100 V, and 200 V applied to the steering electrode. A bias voltage of 1000 V was kept at the cathode in all the measurements. The main amplifier used a 100-ns peaking time. There is a remarkable improvement in the spectral properties of the detector with the application of bias voltages applied to the steering electrode. This effect is due to the collection of electrons from a larger volume and with much less charge sharing.

Despite the fact that all the spectra shown in Figure 3.26 were collected using the same acquisition time, the total number of counts rapidly increased with an increase in the bias voltage applied to the steering electrode. Also, with increasing bias voltage, the majority of counts in the low-energy tail were moved into the photopeak. An energy resolution of 5 keV (FWHM) at 122 keV was measured for the spectrum taken with 200 V applied to the steering electrode (bottom spectrum in Figure 3.26).

The top spectrum in Figure 3.26 models the response of a simple pixel detector with an interpixel pitch of about 0.3 mm and detector thickness of 3 mm. In this case of a simple nondrift structure detector, the small pixels allow for very fast signal formation and the possibility of applying more parallel processing channels per

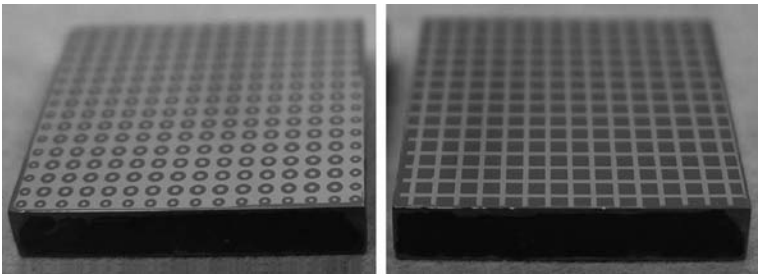


**FIGURE 3.24** Photographs of anode and cathode sides of CZT detector array with parallel drift structures. (© 2009 by IEEE. With permission.)

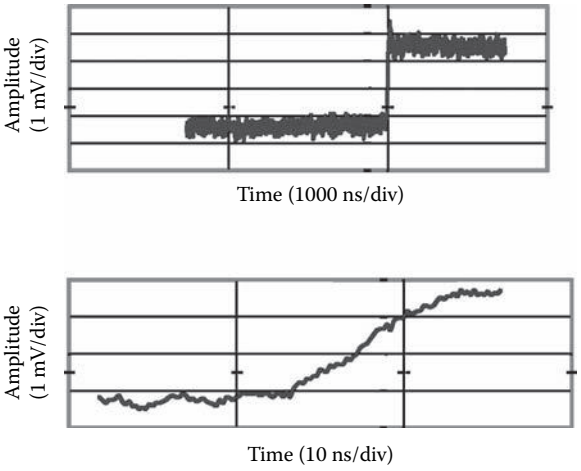



**FIGURE 3.25** A pulse shape from a drift-detector pixel. (© 2009 by IEEE. With permission.)

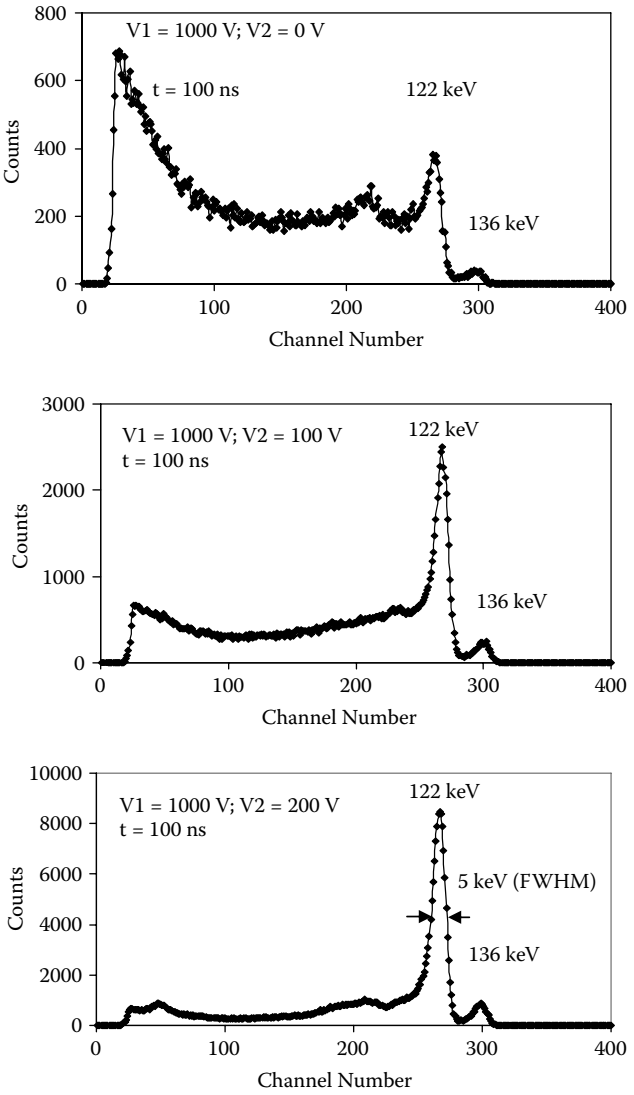



**FIGURE 3.26** Co-57 spectra taken with bias voltage (V2) of 0 V (top spectrum), 100 V (middle spectrum), and 200 V (bottom spectrum) applied to the steering electrode. (© 2009 by IEEE. With permission.)

detector unit area; however, there is a significant distortion in the spectral response. Figure 3.27 presents a pulse from the setup described for Figure 3.26 that was amplified using an Ortec Model 579 fast filter amplifier with an external delay line to shape the output pulse. The length of delay line (coaxial cable) was selected to produce a pulse with a peaking time close to 10 ns.

Pulses from the amplifier were fed into a custom fast multichannel analyzer (MCA.) The resulting spectrum from a Co-57 radiation source is shown in Figure 3.28. An

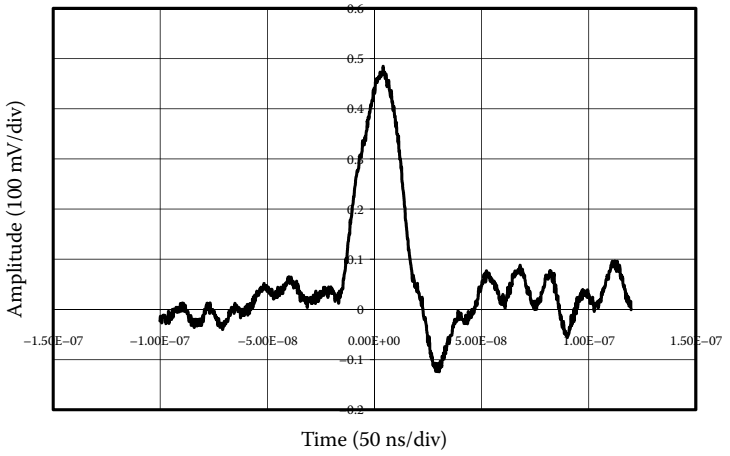



**FIGURE 3.27** A pulse from a drift-detector-array pixel processed by amplifier using a delay line shaper. (© 2009 by IEEE. With permission.)

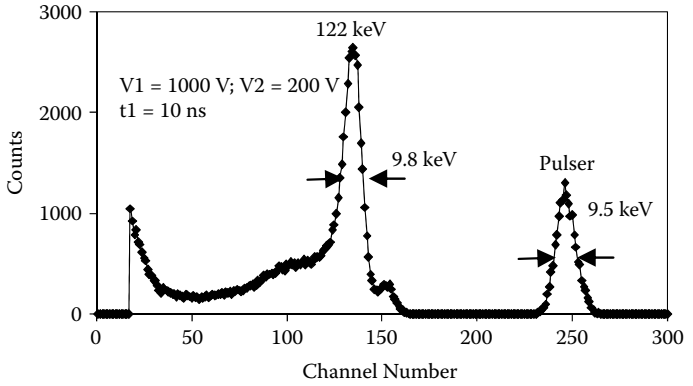



**FIGURE 3.28** Co-57 spectrum taken with a drift-detector-array pixel. (© 2009 by IEEE. With permission.)

energy resolution of 9.8 keV (FWHM) at 122 keV and an electronic noise of 9.5 keV (FWHM) were obtained. An energy resolution of 9.8 keV is sufficient to produce multiple energy images in CT applications. With a 10-ns peaking time (20-ns dead time) for pulses, a paralyzable pulse pileup model allows the calculation of the total throughput (output count rate) of photons. This photon-counting system will be capable of about $15 \times 10^6$ counts per second per pixel at the output, which is nearly three times larger than the output count rate obtained with our previous nondrift structure design, which produced $6 \times 10^6$ counts per second per pixel.

Figure 3.29 shows the output count rate of a typical drift-detector pixel and discrete commercial electronics with pulses formed as shown in Figure 3.27. The output count rate is fairly linear up to $7 \times 10^6$ cps and begins to saturate at about $15 \times 10^6$ cps.

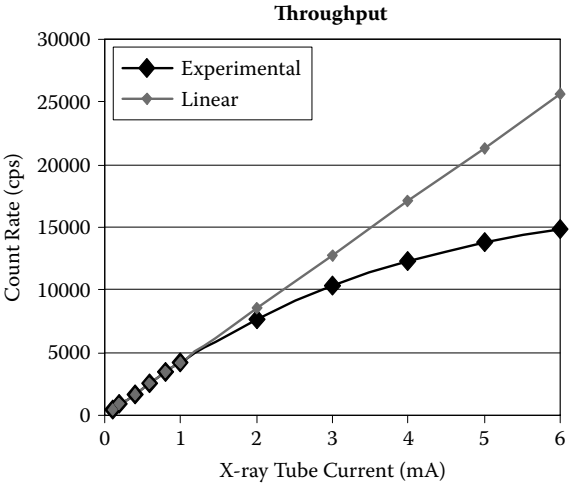



**FIGURE 3.29** Output count rate taken from a single drift-detector pixel versus X-ray tube anode current. A straight line represents a linear fit to data taken at low count rates.

A linear fit to data taken at low count rates shows that the beginning of saturation of the output count rates corresponds to about $25 \times 10^6$ cps at the input. There are several other research groups developing fast photon-counting X-ray imaging detectors; however, to our knowledge, none have to date achieved maximum output count rates as large as those presented here.

## 3.14 X-RAY PHOTON COUNTING WITH SILICON PHOTOMULTIPLIERS

Silicon photomultipliers (SiPMs) or multipixel photon counters (MPPCs) are a new type of photon-counting device. A single SiPM consists of multiple avalanche photodiodes (APDs) connected in parallel and operated in Geiger mode. Each APD subcell of the SiPM generates a pulse signal when it detects one photon. The output signal from the SiPM is the total of the individual signals from all APD subcells. SiPMs offer many excellent performance characteristics compared to conventional photomultiplier tubes (PMTs), semiconductor photodetectors (PDs), and APDs. These performance characteristics include:

- Very fast response time (signal)
- High signal gain
- Single photoelectron response
- Room-temperature operation
- Insensitivity to magnetic fields
- Low-bias voltage operation
- Low power consumption
- Mechanical robustness
- Radiation hardness

These exceptional performance characteristics have caused SiPM devices to be considered a potential replacement for PMTs for many applications, including astronomy, astrophysics, high-energy physics, medicine, biophysics, and national security. The typical front-illuminated SiPM that was used in our experiments has a $1 \times 1$-mm size with 400 subcells (subcell APD areas of $50 \times 50$ μm). The fill factor, a ratio of the active area (light sensitive) of a pixel to the entire area of the pixel, is about 60%. SiPM devices typically consist of 100–5000 subcells located on a common substrate of an area 1–25 $mm^2$ and having a common anode. (The commercially available units are typically about 1 $mm^2$.) Poly-Si resistors are used to decouple individual pixels from each other. The operational bias voltage is in the range 25–80 V at room temperature. The typical gain of a SiPM is $10^5$–$10^6$. For an efficient Geiger (self-sustained) discharge, the applied voltage is 15%–20% higher than the breakdown voltage. An incident photon initiates a Geiger avalanche process that starts to discharge the pixel capacitance until it achieves the voltage level equal to that of the breakdown voltage. The charges produced from individual pixels are summed at a common anode. Because of the identical topology of all pixels, they provide the same gain. The total charge is proportional to the number of pixels discharged. In turn, the output signal is proportional to the number of input photons, assuming that the device operates in the linear dynamic range.

### 3.14.1 Fast Scintillators for Use with Silicon Photomultipliers (SiPMs)

Photon-counting X-ray imaging detectors can be constructed by combining a fast scintillator and an array of SiPMs to maintain the spatial relationship of radiation entering the imaging detector and to provide information about the number of X-ray photons and their energy level. The scintillator consists of numerous segments, each of which is physically aligned with a corresponding SiPM element (pixel) to detect light produced by the interaction of X-rays with that element's scintillator. Optical reflectors between the scintillator segments reflect light back to the segment. We believe that imaging of very high X-ray photon flux utilizing photon counting and energy discrimination with scintillators optically coupled to SiPMs will be possible. This is due to the combination of the very fast decay time and the high light-yield emission of certain scintillators and the highly light-sensitive and extremely fast response of SiPM devices. In recent years, several novel scintillators have been discovered such as LSO, LYSO, YAP, LuAP, and LaBr. These new scintillating materials have characteristics that include

- High light output
- Very fast scintillation decay
- High atomic number (high density)

Table 3.2 summarizes physical properties of the scintillating materials listed here. The requirement for an extremely low afterglow is not an important criterion for the selection of scintillation material in photon-counting CT. In photon-counting mode, the influence of the afterglow on the image can be eliminated by setting the electronic threshold above this signal or above an excess electronic noise due to this effect.

**TABLE 3.2**
**Characteristics of Candidate Scintillator Materials for SiPM-Based X-ray Imaging**

| Scintillator | LSO | LYSO | YAP | LuAP | LaBr |
|---|---|---|---|---|---|
| Density [g/cc] | 7.4 | 7.1 | 5.4 | 8.3 | 5.3 |
| Light yield [photons/keV] | 27 | 32 | 21 | 10 | 61 |
| Effective Z | 66 | 64 | 31.4 | 65 | 46.9 |
| Principal decay time [ns] | 42 | 48 | 25 | 18 | 35 |
| Peak wavelength [nm] | 420 | 420 | 370 | 365 | 358 |
| Index of refraction | 1.82 | 1.8 | 1.94 | 1.95 | 1.88 |
| Hygroscopic | No | No | No | No | Yes |

The peak sensitivity of SiPMs is about 400 nm, which matches reasonably well with all the scintillators listed in Table 3.2. LYSO offers certain advantages over YAP because of its higher light output and better match of the peak wavelength with the peak of the sensitivity of SiPMs. However, YAP has faster decay time. The lower Z of YAP compared to LYSO will still provide sufficient efficiency at CT energies. Also, LSO and LuAP could be scintillators of interest. LaBr, despite having the highest light yield, is highly hygroscopic, which makes it impractical for use in dense segmentation applications.

### 3.14.2 Experimental Results with Silicon Photomultipliers (SiPMs)

Figure 3.30 shows the single-channel LYSO/MPPC X-ray detector that we fabricated. The geometry of the scintillator matches a single segment. We have tested this device with various discrete electronic configurations to determine the maximum output count rate that can be obtained. Figure 3.31 shows the measured output count rate at the counter when the signal is amplified and shaped with an 18-in. delay line shaper connected to an ORTEC 579 amplifier. The maximum output count rate is approximately $16 \times 10^6$ counts per second per pixel. The high output rate is due to the fast signal response time and use of a fast scintillator (see Table 3.2).

The results of an experiment with $^{241}$Am and $^{137}$Cs sources are shown in Figure 3.32. The spectrum on the left clearly shows the 60-keV gamma-ray photopeak from the $^{241}$Am source. The spectrum on the right shows the 32-keV Ba K X-ray photopeak from the $^{137}$Cs source. The separation of the Ba K peak from the electronic-noise shoulder indicates a very low electronic-noise level of about 7 keV. Despite this low electronic-noise level, the energy resolution is much worse than that possible with semiconductor detectors such as CdTe or CZT. The lower energy resolution is due to the basic phenomena related to the mechanisms that generate the signal in the scintillator and the multiplication noise in the SiPM.

## 3.15 SUMMARY

The chapter presents novel detector structures that have been specifically developed or evaluated for photon-counting X-ray imaging applications. Most of our work was

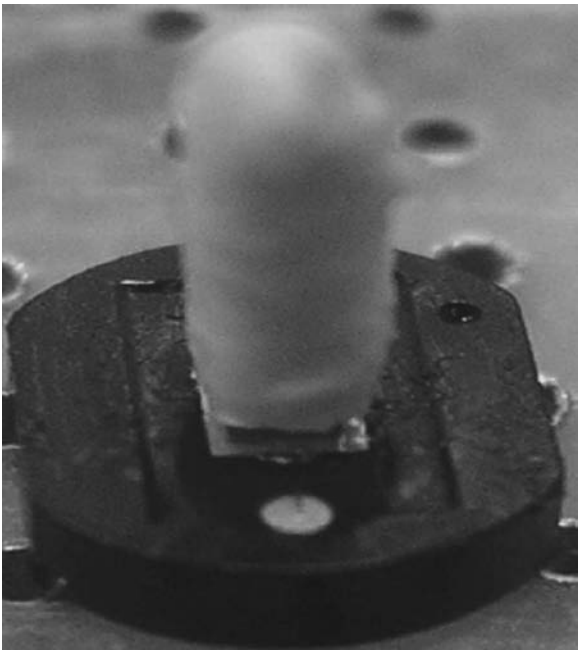


**FIGURE 3.30** A 1-mm$^2$ SiPM (Hamamatsu Model S10361-11-050C) light detector coupled to a $1 \times 1 \times 5$-mm LYSO scintillator.

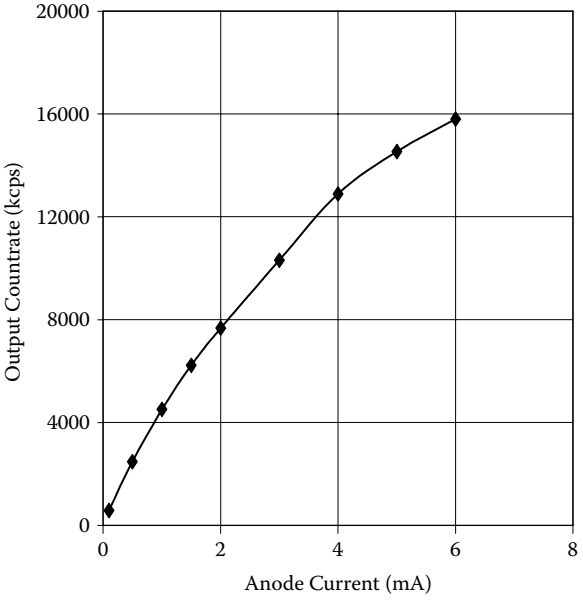



**FIGURE 3.31** A 1-mm$^2$ SiPM light detector coupled to a $1 \times 1 \times 5$-mm LYSO scintillator with an 18-in.-long delay line shaper.

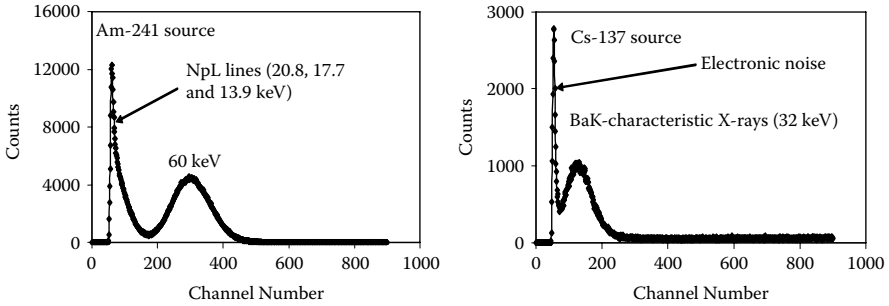



**FIGURE 3.32** $^{241}$Am (left side) and $^{137}$Cs (right side) spectra from the 1-mm$^2$ SiPM light detector coupled to a $1 \times 1 \times 5$-mm LYSO scintillator.

concentrated on development of CdTe or CZT detector arrays; on the use of fast, low-noise ASIC readout electronics; and on solving the problems of interconnecting the detectors and electronics to create modules that can be tiled together. The developed detector structures based on compound semiconductor detectors provide a signal response much faster than the transit time of carriers over the whole detector thickness. At the same time, the individual detector pixels exhibit good spectral performance at very short peaking times.

We have developed a fast and low-noise ASIC electronics readout that is optimized for amplification and processing of signals from CZT and CdTe detector arrays. Detector modules containing crystals and ASIC electronics were specifically designed to fit into a commercial scanner and replace the conventional integrating scintillator detectors with fast and highly efficient photon-counting and energy-dispersive detectors. The output count rate of $6 \times 10^6$ cps per 1-mm$^2$ pixel in a fully functional CT detector module has been demonstrated. The first patient CT images were taken with these photon-counting, energy-dispersive detectors. Despite a relatively low X-ray flux, the images demonstrate good image quality, showing great promise for possibly decreasing the radiation dose to patients while also providing tissue decomposition capabilities. Detector arrays with parallel drift-detector structures, when optimized and combined with fast ASIC electronics, can provide a good developmental path for photon counting at the very high X-ray flux (≈100 Mphotons/mm$^2$/s) that is currently used in typical CT exams with conventional integrating detectors.

We also investigated commercial SiPM (MPPC) devices coupled to a fast scintillator (LYSO) for fast X-ray photon counting. The output count rate of $15 \times 10^6$ cps per 1-mm$^2$ pixel has been achieved. However, energy resolution measured with these devices was much worse than that possible with semiconductor detectors such as CdTe or CZT. The higher resolution of CdTe and CZT is due to more efficient generation of the signal in the semiconductors and the use of conventional electronic amplification.

## ACKNOWLEDGMENTS

The authors would like to thank Peter Weilhammer, Danielle Moraes, Pierre Jarron, and Jan Kaplon of CERN for their assistance in the development of the ASIC. We

also thank Dr. Oded Meirav and Jerry Arrenson of GE Healthcare, Ltd., Haifa, Israel, for providing images of patients. Dx Ray acknowledges support for this work was received from NIH Grants 1R43RR024101, 2R44RR024101, and 1R44EB008612 and the BIRD Foundation.

## REFERENCES

1. Hu, H., H. D. He, W. D. Foley, and S. H. Fox. 2000. Four multidetector-row helical CT: Image quality and volume coverage speed. *Radiology* 215:55–62.
2. Whiting, B. R., P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson. 2006. Properties of preprocessed sinogram data in X-ray computed tomography. *Medical Physics* 33:3290–3303.
3. Swank, R. K. 1973. Absorption and noise in X-ray phosphors. *J. Appl. Phys.* 44:4199–4203.
4. Mukundan, S., K. B. Ghaghada, C. T. Badea, C. Y. Kao, L. W. Hedlund, J. M. Provenzale, G. A. Johnson, E. Chen, R. V. Bellamkonda, and A. Annapragada. 2006. A liposomal nanoscale contrast agent for preclinical CT in mice. *Am. J. Roentgenol*. 186:300–307.
5. Mettler, F. J., P. Wiest, J. Locken, and C. Kelsey. 2000. CT scanning: Patterns of use and dose. *J. Radiol. Prot*. 20:353–359.
6. Patterson, S.A., H. I. Khalil, and D. M. Panicek. 2006. MRI evaluation of small hepatic lesions in women with breast cancer. *Am. J. Roentgenol*. 187:307–312.
7. Centers for Disease Control and Prevention. 2006. Statistics: Lung cancer. http://www.cdc.gov/lungcancer/statistics/index.htm.
8. American Cancer Society. 2006. Cancer statistics 2006 presentation. http://www.cancer.org/docroot/PRO/content/PRO_1_1_Cancer_Statistics_2006_Presentation.asp.
9. Greenlee, R. T., T. Murray, S. Bolden, and P. A. Wingo. 2000. Cancer statistics 2000. *CA Cancer J. Clin*. 50:7–33.
10. Landis, S. H., T. Murray, S. Bolden, and P. A. Wingo. 1999. Cancer statistics 1999. *CA Cancer J. Clin*. 49:8–31.
11. Landis, S. H., T. Murray, S. Bolden, and P. A. Wingo. 1998. Cancer statistics 1998. *CA Cancer J. Clin*. 48:6–29.
12. Swensen, S. J., J. R. Jett, and T. E. Hartman. 2003. Lung cancer screening with CT, Mayo Clinic experience. *Radiology* 226:756–761.
13. Sobue, T., N. Moriyama, and M. Kaneko. 2002. Screening for lung cancer with low-dose helical computer tomography: Anti-lung cancer association project. *J. Clin. Oncol*. 20:911–920.
14. Gary, K., R. L. Keith, and T. Byers. 2002. Randomized controlled trial with low-dose spiral CT for lung cancer screening: Feasibility study and preliminary results. *Radiology* 225:506–510.
15. Nawa, T., T. Nakagawa, S. Kusano, Y. Kawasaki, Y. Sugawara, and H. Nakata. 2002. Lung cancer screening using low dose spiral CT: Results of baseline and 1-year follow-up studies. *Chest* 122:15–20.
16. Henschke, C. I., D. I. McCauley, D. F. Yankelevitz, D. Naidich, G. McGuinness, and O. Miettinen. 1999. Early Lung Cancer Action Project: Overall design and findings from baseline screening. *Lancet* 354:99–105.
17. Brenner, D. J. 2004. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. *Radiology* 231:440–445.
18. Hillman, B. J. 2003. Economic, legal, and ethical rationales for the ACRIN national lung screening trial of CT screening of lung cancer. *Acad. Radiol*. 10:349–350.
19. Silverberg, E., and J. A. Libera. 1989. Cancer statistics 1989. *CA Cancer J. Clin*. 39:3–20.

20. Thoeni, R. F., and I. Laufer. 1994. Polyps and cancer. *Gastrointestinal Radiology* 1:1160–1199.
21. Hara, A. K., C. D. Johnson, J. E. Reed, D. A. Ahlquist, H. Nelson, R. L. MacCarty, W. S. Harmsen, and D. M. Ilstrup. 1997. Detection of colorectal polyps with CT colography: Initial assessment of sensitivity and specificity. *Radiology* 205:59–65.
22. Kalendar, W. A. 1986. Evaluation of a prototype dual-energy computed tomographic apparatus: Phantom Studies. *Medical Physics* 13:334–339.
23. Johns, P. C., J. Dubeau, D. G. Gobbi, M. Li, and S. Dixit. 2002. Photon-counting detectors for digital radiography and X-ray computed tomography. *SPIE* TD01:367–369.
24. Shikhaliev, P. M., X. Tong, and S. Molloi. 2005. Photon computed tomography: Concept and initial results. *Med. Phys*. 32:427–436.
25. Maidment, A., C. Ullberg, K. Lindman, L. Adelow, J. Egerstrom, M. Eklund, T. Francke, U. Jordung, T. Kristoffersson, L. Lindqvist, D. Marchal, H. Olla, E. Penton, J. Rantanen, S. Solokov, N. Weber, and H. Westerberg. 2005. Evaluation of a photon-counting breast tomosynthesis imaging system. *Medical Imaging 2005, Proceedings of the SPIE*. 5745:572–582.
26. Lundqvist, M., B. Cederstrom, V. Chmill, M. Danielsson, and B. Hasegawa. 2001. Evaluation of a photon-counting X-ray imaging system. *IEEE Trans. Nucl. Sci.* 48:1530–1536.
27. Michel, T., G. Anton, M. Bohnel, J. Durst, M. Firsching, A. Korn, B. Kreisler, A. Loehr, F. Nachtrab, D. Niederlohner, F. Sukowski, and P. T. Talla, "A fundamental method to determine the signal-to-noise ratio (SNR) and detective quantum efficiency (DQE) for a photon counting pixel detector," *NIM A*, vol. 568(2), pp. 799–802, 2006.
28. Danielsson, M., and H. Bornefalk. 2001. Dose-efficient system or digital mammography. *IEEE Trans. Nucl. Sci*. 48:1530–1536.
29. Sangwaiya, M. J., M. K. Kalra, A. Sharma, E. F. Halpern, J. Shepard, and S. R. Digumarthy, "Dual-energy computed tomographic pumonary angiography: a pilot study to assess the effect of image quality and diagnostic confidence," *Journal of Computer Assisted Tomography*, vol. 34, pp. 46–51, 2010.
30. Tapiovaara, M. J., and R. Wagner. 1985. SNR and DQE analysis of broad spectrum X-ray imaging. *Physics in Medicine and Biology* 30:519–529.
31. Shikhaliev, P. M. 2005. Beam hardening artefacts in computed tomography with photon counting, charge integrating and energy weighting detectors: A simulation study. *Physics in Medicine and Biology* 50:5813–5827.
32. Dilmanian, F. A. 1995. Dual energy iodine contrast CT with monochromatic X-rays. *Nuclear Science Symposium and Medical Imaging Conference, IEEE* 3:1392–1396.
33. Bhalla, M. 1995. Dual kVp CT to detect calcification in solitary pulmonary nodule. *Journal of Computer Assisted Tomography* 1(4): 44–47.
34. Vetter, J. R., W. H. Perman, W. A. Kalender, R. B. Mazess, and J. E. Holden. 1986. Evaluation of a prototype dual-energy computed tomographic apparatus: II, Determination of vertebral bone mineral content. *Medical Physics* 13:340–343.
35. Alvarez, R.E., and A. Macovski. 1976. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine and Biology* 21:733–744.
36. Brody, W. R., G. Butt, A. Hall, and A. Macovski. 1981. A method for selective tissue and bone visualization using dual energy scanned projection radiography. *Medical Physics* 8:353–357.
37. Lehmann, L. A., R. E. Alvarez, A. Macovski, W. R. Brody, N. J. Pelc, S. J. Riederer, et al. 1981. Generalized image combinations in dual KVP digital radiography. *Medical Physics* 8:659–667.
38. Vetter, J. R., and J. E. Holden. 1988. Correction for scattered radiation and other background signals in dual-energy computed tomography material thickness measurements. *Medical Physics* 15:726–731.

39. Engler, P., and W. D. Friedman. 1990. Review of dual-energy computed tomography techniques. *Materials Evaluations* 48:623–629.
40. Kirby, B. J., J. R. Davis, J. A. Grant, and M. J. Morgan. 2003. Extracting material parameters from X-ray attenuation: A CT feasibility study using kilovoltage synchrotron X-rays incident upon low atomic number absorbers. *Physics in Medicine and Biology* 48:3389–3409.
41. Drost, D. J., and A. Fenster. 1980. Experimental dual xenon detectors for quantitative CT and spectral artifact correction. *Medical Physics* 7:101–107.
42. Torikoshi, M., T. Tsunoo, M. Endo, M. Natsuhori, T. Kakizaki, N. Ito, et al. 2005. Features of dual-energy X-ray computed tomography. *Nuclear Instruments and Methods in Physics Research A* 548:99–105.
43. Kappler, S., and S. Wirth. 2008. Comparison of dual-kVp and dual-layer CT in simulations and real CT system measurements. *Nuclear Science Symposium and Medical Imaging Conference, IEEE* 4828–4831.
44. Chmeissani, M., and C. Frojdh. 2004. First experimental tests with a CdTe photon counting pixel detector hybridized with a Medipix2 readout chip. *IEEE Trans. Nucl. Sci*. 51:2379–2385.
45. Heismann, B. J. 2006. Signal-to-noise Monte Carlo analysis of base material decomposed CT projections. *Medical Imaging 2006: Physics of Medical Imaging, San Diego, SPIE* 214–237.
46. Bisogni, M., and M. Gambaccini. 2002. A digital system based on a bichromatic X-ray source and a single photon counting device: A single-exposure dual-energy mammography approach. *Proc. SPIE* 4682:620–632.
47. Mikulec, B., M. Campbell, E. Heijne, X. Llopart, and L. Tlustos. 2003. X-ray imaging using single photon processing with semiconductor pixel detectors. *NIM A* 511:282–286.
48. Cahn, R. N., B. Cederstrom, M. Danielsson, A. Hall, M. Lundqvist, and D. Nygren. 1999. Detective quantum efficiency dependence on X-ray energy weighting in mammography. *Med. Phys.* 26:2680–2683.
49. Szeles, C., S. A. Soldner, S. Vydrin, J. Graves, and D. S. Bale. 2007. Ultra high flux 2-D CdZnTe monolithic detector arrays for X-ray imaging application. *IEEE Trans. Nucl. Sci*. 54:1350–1358.
50. Bennett, P. R., K. S. Shah, L. J. Cirignano, L. P. Klugerman, L. P. Moy, and M. R. Squillante. 1999. Characterization of polycrystalline TlBr films for radiographic detectors. *IEEE Trans. Nucl. Sci*. 46:266–270.
51. Shah, K. S., P. Bennett, Y. Dimitriyev, L. Cirignanno, M. Klugerman, M. R. Squillante, R. A. Street, J. T. Rahn, and S. E. Ready. 1999. PbI2 for high resolution digital X-ray imaging. *Proceedings of the SPIE* 3770:164–173.
52. Street, R. A., J. T. Rahn, S. E. Ready, K. Shah, P. R. Bennett, Y. Dimitriyev, P. Mei, J. P. Lu, R. B. Apte, J. Ho, K. Schylenbergh, F. Lemmi, M. Braiman, and P. Nylen. 1999. X-ray imaging using lead iodide as a semiconductor detector. *Proceedings of the SPIE* 3659:36–47.
53. Schieber, M., H. Hermon, A. Zuck, A. Vilensky, L. Melekhov, R. Shatunovsky, E. Meerson, and H. Saado. 1999. Polycrystalline mercuric iodide detectors. *Proceedings of the SPIE* 3770:146–155.
54. Schieber, M., A. Zuck, L. Melekhov, R. Shatunovsky, H. Hermon, and R. Turchetta. 1999. High flux X-ray response of composite mercuric iodide detectors. *Proceedings of the SPIE* 3768:296–309.
55. Yoon, H., M. Van Scyoc, M. Goorsky, H. Hermon, M. Schieber, J. Lund, and R. James. 1997. Investigation of the effects of polishing and etching on the quality of Cd1-xZnxTe using spatial mapping techniques. *J. Electronic Materials* 26:529–533.

56. Rouse, A. A., C. Szeles, J. O. Ndap, S. A. Soldner, K. B. Parnham, D. J. Gaspar, M. H. Engelhard, A. S. Lea, S. V. Shutthanandan, T. S. Thevuthasan, and D. R. Baer. 2002. Interfacial chemistry and the performance of bromine-etched CdZnTe radiation detector devices. *IEEE Trans. Nucl. Sci*. 49:2005–2009.
57. Mescher, M., J. Hoberg, T. Schlesinger, and R. James. 1999. Analysis on the measurement of leakage currents in CdZnTe radiation detectors. *IEEE Trans. Nucl. Sci*. 46:2289–2296.
58. Werthen, J., J. P. Haring, A. Fahrenbruch, and R. Bube. 1983. Effects of surface preparation on the properties of metal/CdTe junctions. *J. Appl. Phys*. 54:5982–5989.
59. Knoll, G. F. 1999. *Radiation detection and measurement*. 3rd ed., 792–793. New York: Wiley.
60. Barber, W. C., N. Malakhov, M. Q. Damron, N. E. Hartsough, D. Moraes, P. Weilhammer, E. Nygard, and J. S. Iwanczyk. 2006. Guard ring elimination in CdTe and CdZnTe detectors. *IEEE Nucl. Sci. Symp. Conference Record* 2414–2416.
61. Arodzero, A., W. C. Barber, M. Q. Damron, N. E. Hartsough, J. S. Iwanczyk, N. Malakhov, E. Nygard, D. Moraes, P. Weilhammer, and P. Jarron. 2006. A system for the characterization and testing of CdZnTe/CdTe pixel detectors for X-ray and gamma-ray imaging. *IEEE Nucl. Sci. Symp. Conference Record* 3638–3642.
62. Jarron, P., F. Anghinolfi, E. Delagne, W. Dabrowski, and L. Scharfetter. 1996. A transimpedance amplifier using a novel current mode feedback loop. *Nuclear Instruments and Methods in Physics Research A* 377:435–439.
63. Kaplon, J., and W. Dabrowski. 2005. Fast CMOS binary front-end for silicon strip detectors at LHC experiments. *IEEE Trans. Nucl. Sci*. 52:2713–2720.
64. Moraes, D., E. Nygard, and A. Rudge. 2005. Front-end electronics for the readout of CdZnTe sensors. *Nuclear Science Symposium Conference Record* 1:348–352.
65. Wu, X. J. K. Brown, K. Kalki, and B. H. Hasegawa. 1996. Characterization and correction of pile-up in simultaneous emission-transmission computed tomography. *Med. Phys*. 23:569–575.
66. Gui-Nian, D., K. E. Turner. 1989. A method of correcting dead time and pulse pile-up errors in energy-dispersive X-ray analysis. *X-ray Spectrometry* 18:57–62.
67. Iwanczyk, J. S., E. Nygard, O. Meirav, J. Arenson, W. C. Barber, N. E. Hartsough, N. Malakhov, and W. J. C. 2007. Photon counting energy dispersive detector arrays for X-ray imaging. IEEE Nuclear Science Symposium and Medical Imaging Conference. 1st ed., vol. M09-04, Honolulu, 2741–2748.
68. Iwanczyk, J. S., E. Nygard, J. Meirav, J. Arenson, W. C. Barber, N. E. Hartsough, N. Malakhov, and J. C. Wessel, "Photon counting energy dispersive detector arrays for x-ray imaging," *IEEE Trans. Nucl. Sci.*, vol. 56, pp. 535–542, 2009.

# 4 Planar and PET Systems for Drug Development

*Ryoko Yamada and Hiroshi Uchida*

**CONTENTS**

## 4.1 INTRODUCTION

A radiation detector has been developed to be applied for nuclear imaging in nuclear medicine. According to Townsend's latest review [1], nuclear imaging technology has been migrating into the field of nuclear medicine since the late 1940s. Today, the tomographic imaging of the human body using radioisotopes, such as positron emission tomography (PET), single-photon-emission computed tomography (SPECT), etc., is indispensable for visualizing the functions of biological mechanisms noninvasively, e.g., metabolism or neurotransmitter. The progress of tomographic imaging tools relates strongly to the progress that has been made in scintillation detector technology, involving both advances of photodetector technology and the discovery of novel materials for scintillators. Recently, the advent of a hybrid system of PET/SPECT and X-ray computed tomography (CT) achieved precise superimposition of the functional image on the anatomical image. This property of the hybrid system brought benefits to medical diagnosis, such as the whole-body health examination for cancer detection, and contributed significantly to the spread of nuclear imaging apparatus into clinical use.

The visualization technique using this hybrid system is called *hardware fusion* in contrast to *software fusion*. The hardware fusion makes it possible to superimpose the functional image acquired by nuclear imaging modality (PET, SPECT, etc.) on the anatomical image acquired by radiological imaging modality (X-ray CT). Both images are obtained under the same conditions, such as scanning axis, geometrical configuration of target, and so on. Besides the improvement in diagnostic accuracy, it becomes possible to avoid the need for an external sealed radioisotope, which saves cost and shortens the measurement time by utilizing the anatomical image for the attenuation correction.

Further sophistication of human whole-body PET has been obtained by using time of flight (TOF) information with fast scintillation detectors to secure improved signal-to-noise ratio in the image. Such sophistication requires developers of detector systems to design their products considering the whole measurement system and how it is applied. Moreover, the requirement for progress in the development of hardware arises from expansion in the field of applications. Above all, nuclear imaging is expanding into molecular imaging, based on the knowledge of the specific genes and proteins involved in disease mechanisms. Nowadays, the molecular imaging requires not only nuclear-imaging methods, but also other medical noninvasive imaging methods, such as radiological imaging, including X-ray CT and magnetic resonance (MR), optical imaging with either fluorescence and/or bioluminescence, and ultrasound imaging, with new contrast agents. One can find detailed explanations [2, 3] on how these clinically employed imaging methods are utilized to visualize specific molecules and on the significance of the molecular imaging applications.

Among the various applications of molecular imaging, the development of new pharmaceuticals should be the most significant. From this viewpoint, the demand on imaging systems dedicated to in vivo animal studies is aiming for (a) preclinical evaluation of new pharmaceuticals using small laboratory animals and (b) research on the fundamental mechanisms of human diseases using new animal models that

mimic human disorders. Besides choosing either of these two approaches, the development of target-specific molecular probes is considered to be most important. Nevertheless, continued progress in the development of new pharmaceuticals will not be possible unless the improvements in the development of detectors can keep up with the advances of the target-specific molecular probes, as these detectors must be able to visualize faint signals emitted from the probes. For this purpose, systems must be developed to provide dynamic and high-throughput measurements with sufficiently high spatial resolution, sensitivity, and quantitative capability. However, some of the specifications conflict with each other, making it difficult to develop a system satisfying all requirements.

In an effort to resolve the problem, we divided those requirements into two groups and developed systems, each of which satisfies either of the requirements: One is a small-animal PET scanner, and the other is a planar positron imaging system (PPIS).

In this chapter, we describe in detail the total construction and fundamental specifications of PPIS [4] and PET [5] systems. Both systems are based on a common physical principle to detect a pair of almost simultaneously arriving gamma rays emitted in the opposite direction from electron–positron pair annihilation process, although the geometrical configurations of the systems and their derivative characteristics are very different.

## 4.2 PPIS

### 4.2.1 Introduction

Although the usefulness of PET is recognized, some animal experiments are incapable of obtaining high-quality images because of the large statistical noise. This problem is particularly serious when the specific activity of the radiotracer is low, or when it is necessary to measure a series of short-period images for pharmacokinetic analysis of new drugs. In this kind of evaluation, an imaging system is often required to measure a large number of samples efficiently. Unfortunately, the small-animal PET scanners available today have not been designed for high-throughput measurements.

Thus, a high-resolution planar imaging system was developed for small-animal experiments to evaluate new pharmaceuticals. Planar coincidence detectors have a long history, having been used in the early stage of PET systems development [6–8], and they have recently been used in PET/SPECT hybrid systems [9–14]. However, the new planar imaging system was designed to provide only focal-plane images instead of the transaxial images provided with PET scanners. Simplicity in the system design includes several benefits such as high resolution, low cost, easy operation, and a high-throughput measurement capability. Also, it is suitable for measuring a subject in a short period of time or with very-low-activity tracers. Recent years have seen a surge in interest in imaging the function of biological phenomena occurring in living creatures such as plants, and studies with similar planar imaging systems will likely increase [15–22].

### 4.2.2 System Construction and Data Processing

The planar imaging system is designed to detect radiation from a subject with two opposing detector heads. Figure 4.1 shows the construction of the detector head. To achieve sufficient spatial resolution, the detector heads have the following three-fold hierarchical structure. Each detector head contains 2 (columns) × 3 (rows) of detector modules, where each detector module is composed of 2 × 2 block-detector units, each of which consists of 10 × 10 BGO (bismuth germanate) crystal arrays optically coupled to a metal-packaged PS-PMT (position-sensitive photomultiplier tube) (Hamamatsu R8520-00-C12) [23]. In the following discussion, we describe in detail the detector and circuit configuration in PPIS (Hamamatsu PPIS-4800).

#### 4.2.2.1 Detector Module

Figure 4.2 illustrates the construction of the detector module and its readout circuit, along with a schematic drawing of the detector unit. The BGO crystal elements (dimensions of each crystal element are 2 × 2 mm in cross-section and 20 mm in length) are optically isolated from each other with 0.2-mm-thick Teflon sheets to avoid inter-element cross talks. Once gamma rays from a subject strike a BGO crystal,

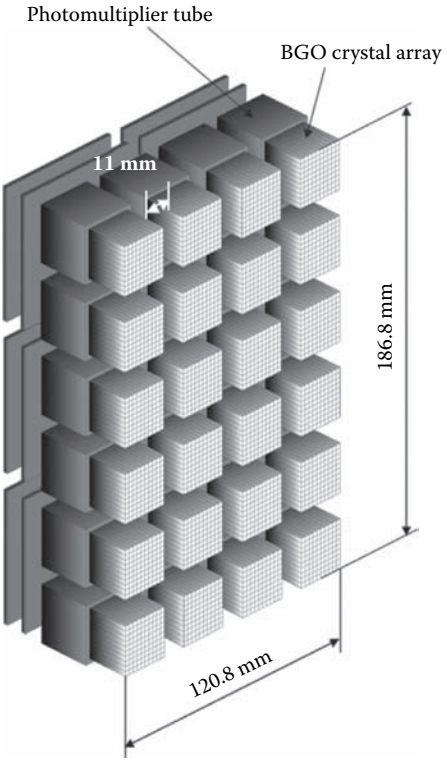



**FIGURE 4.1** Schematic drawing of the detector head consisting of 2 (columns) × 3 (rows) detector modules. This detector module is composed of 2 (columns) × 2 (rows) block-detector units, with each unit having a BGO scintillation array.

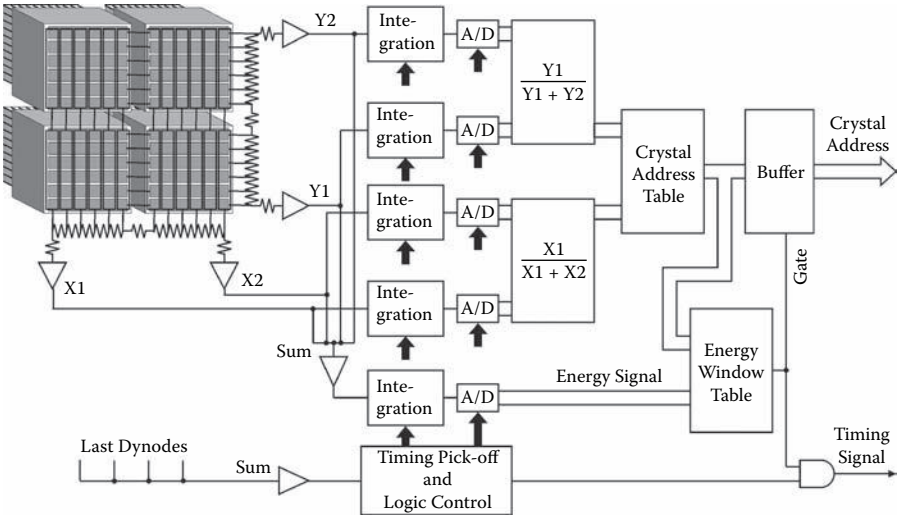



**FIGURE 4.2** Block diagram of the detector-module readout system. (© 2004 by Elsevier B.V. With permission.)

they interact with heavy atoms to stochastically produce scintillation photons. The scintillation photons travel in the crystal element, undergoing multiple reflections at the element's walls, and finally arrive at the photocathode of the PS-PMT. Here we describe in detail what occurs in the PS-PMTs. The incident photons are converted to photoelectrons at the photocathode [24], multiplied by $5 \times 10^5$–$10^6$ through channel-separated dynodes keeping positional information, and detected by $6(X) + 6(Y)$ crossed-plate anodes.

We can obtain information on the total energy of incident radiation and the position from output signals of the PS-PMT as follows. The total energy of incident radiation is proportional to the total charge amount of the PS-PMT output. On the other hand, the position of an incident radiation is calculated as the centroid of the charge distribution on the plate anodes. To obtain the centroid information, the *X*- and *Y*-plate anodes of four PS-PMTs in the same module are connected to the common *X* and *Y* resistor-chains, respectively. The position signals in the *X* direction ($X_1$, $X_2$) and those in the *Y* direction ($Y_1$, $Y_2$) are amplified and fed into a signal-processing unit. It is also noted that the outputs from the final dynodes of four PS-PMTs in one detector module are all summed to provide the timing information of incident photons (Figure 4.2).

#### 4.2.2.2 Detector Head

The block-detector units in a detector head are arranged on a plane with a gap of 11.0 mm between crystal arrays, which is equal to the width of five crystal elements. This arrangement of the block-detector unit with gaps enhances the effective field-of-view (FOV) area without generating any dead area on the focal imaging planes. The FOV area of $120.8 \times 186.8$ mm is obtained with a detector area of $80 \times 120$ mm. Using this setup, we detect coincidence events between one module

and the $3 \times 3$ array of modules directly opposite it, assuming that the radiations from objects arrive at the corresponding pair of detector modules almost simultaneously. The dimension of the detector head is only 225 mm (height) $\times$ 180 mm (width) $\times$ 115 mm (length). Such a compact and simple structure of the detector head allows very versatile arrangement of detectors. For example, Figure 4.3 illustrates the horizontal- and vertical-mode positioning of the detectors, and Figure 4.4 is a photograph of the PPIS.

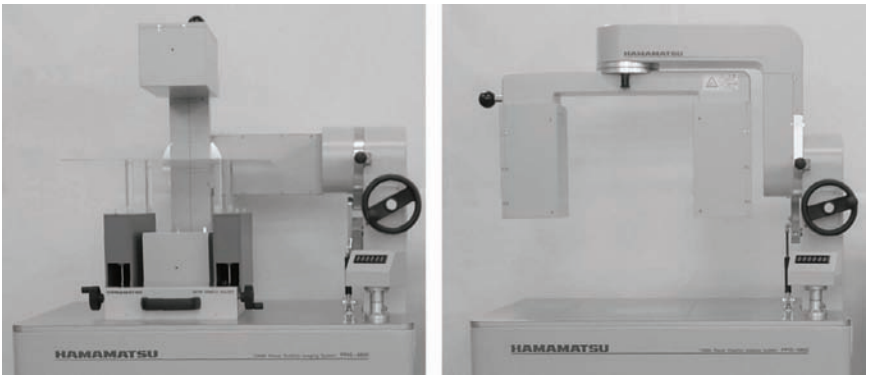



**FIGURE 4.3** Photograph of the planar positron-imaging device. Left: Horizontal positioning of detectors. Right: Vertical positioning of detectors.


**FIGURE 4.4** Photograph of PPIS.

In our setup, detector heads are not particularly shielded from external radiation. However, the side surfaces of each crystal array are covered with 1-mm-thick lead plates to suppress the oversensitivity of peripheral crystal elements to gamma rays directly penetrating their sides.

#### 4.2.2.3 Electronics

The position signals ($X_1$, $X_2$, $Y_1$, $Y_2$) and the timing signal from each detector module are fed to a signal-processing circuit, and the four position signals are summed to produce an energy signal. The position and energy pulses, which were shortened with 200-ns delay line clipping, are integrated with gated-integrator circuits and converted to 8-bit digital signals. The positions in the $X$ and $Y$ directions are calculated from the digitized position signals with the help of position lookup tables (LUT) stored in read-only memory (ROM), and then converted to a crystal address code with the position-to-address conversion table in a flash ROM. The energy signal and the crystal address code are transferred to a flash ROM, where the energy windows are set corresponding to each crystal. When the energy signals match the preset energy windows, the timing and crystal codes are transferred to the following coincidence-detection circuits (Figure 4.2).

#### 4.2.2.4 Data Processing and Image Reconstruction

Coincidence is detected when two timing signals arrive at the coincidence circuit within an interval of 20 ns. A pair of crystal codes detected as a coincidence event are acquired by a personal computer (PC) through a small-computer system interface (SCSI). The acquired data is stored in a histogram memory during the preset frame duration and then stored onto a hard disk (HD), where two types of data storage modes can be selected. The acquired crystal pair codes are directly stored in the histogram memory, or the crystal pair codes are accumulated in the histogram memory after being converted to the coordinate ($X$, $Y$) data on a focal plane. The stored data on the HD is then transferred to another PC through an Ethernet connection for image processing and data analysis.

The position ($X_0$, $Y_0$) of a positron annihilation in an observed object is determined as an intersection of the focal plane with the line connecting detection points ($X_a$, $Y_a$) and ($X_b$, $Y_b$) on a pair of directly facing detectors (i.e., coincidence line). A coincidence line is back-projected to be assigned a $1.1 \times 1.1$-mm square pixel on the focal plane. We can obtain a cross-sectional image as a density distribution of the back-projected pixels through the calibration based on the sensitivity variations of the detectors, which are determined from an image for the flat phantom, as described later.

The acceptance angle of the coincidence lines determines the spatial resolution of the system, i.e., a detection with a small acceptance angle improves the spatial resolution at the cost of detection efficiency. The present system has two mode-of-acceptance angles, which are called the wide mode and the narrow mode. The acceptance angle in the wide mode includes the coincidence events detected between one module and the $3 \times 3$ array of modules directly opposite it, and this is the maximum angle, as limited by the hardware. The acceptance angle in the narrow mode is limited to the coincidence events detected between one block-detector unit and the $3 \times 3$ array of

block-detector units directly opposite it, and is limited by the software. The narrow mode is used for evaluation of system performance and applicability, as discussed in the following section.

The focal-plane imaging is free from noise amplification in the tomographic reconstruction process; hence, it provides plane images with a superior signal-to-noise ratio, even with a short period of data acquisition, and even for objects involving low-activity radiotracers. Finally, note that care is required when observing a thick object with the planar system. In this case, activities existing outside the focal plane are projected onto the focal-plane image as a blurred distribution to increase the background level on the focal-plane image, resulting in the degradation of contrast.

### 4.2.3 System Performance

Here we present the results of evaluating the system performance of PPIS.

#### 4.2.3.1 Spatial Resolution

The spatial resolution for a point source on the focal plane is determined from the coincidence response functions (CRFs) of all possible crystal pairs back-projected onto the focal plane. The CRF for the gamma rays obliquely penetrating each crystal spreads widely as the angle of the coincidence line increases, leading to the degradation of spatial resolution or so-called parallax error. Thus, the spatial resolution depends on the distance between a pair of opposing detector heads (referred to as the D-D distance in the following) and the acceptance angle. The spatial resolution also varies from position to position on the focal plane because of the geometrical arrangement of the crystals.

Next, we evaluated the relationship between the spatial resolution and the D-D distance. The measurement of the spatial resolutions was performed with a $^{22}$Na point source (1-mm diameter) placed at three different positions on the focal plane as a function of the D-D distance (Figure 4.5). Here, the acceptance angle was selected according to the narrow-mode operation. The measured spatial resolution was compared with that obtained by the Monte Carlo simulations. In this simulation, the distribution of energy deposits in the scintillator was calculated for two gamma rays emitted from a 1-mm-diameter spherical source while considering the angular deviation of the two annihilation gamma rays from 180 degrees.

Figure 4.5 shows the measured and simulated spatial resolutions at different positions on the focal plane and for different D-D distances. The discrepancy in the spatial resolution between the measured and simulated data was mainly due to the spatial resolution of the PS-PMT, as well as partly due to the effect of positron range. These phenomena were disregarded in the simulation, which indicated that the parallax error decreased as the D-D distance increased, and that the effect of angular deviation became dominant for D-D distances over 30 cm. Additionally, the dependence of the spatial resolution on the source position decreased as the D-D distance increased. From these results, we fixed the D-D distance to 30 cm in the following measurements.

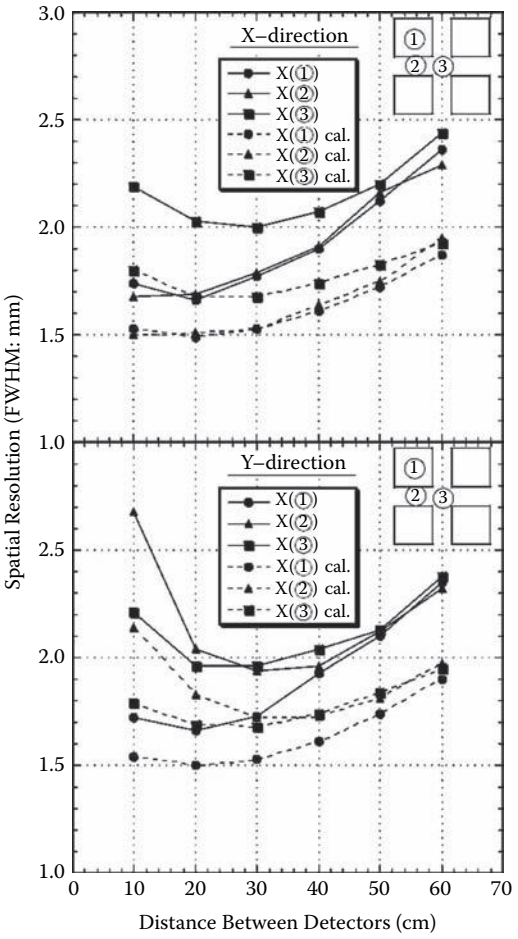



**FIGURE 4.5** Dependence of spatial resolution on the distance between detectors (D-D). Measured spatial-resolution values (FWHM, mm) are shown with solid lines, and those by simulation are shown with broken lines. Labels 1, 2, and 3 indicate the position of the $^{22}$Na point source on the focal plane (mid-plane). (© 2004 by Elsevier B.V. With permission.)

At a D-D distance of 30 cm, we measured the variations of the spatial resolutions on two column lines by scanning the point source with 5.5-mm steps, the results of which are shown in Figure 4.6. The spatial resolution in column B (on the line above the center of crystal blocks) was slightly improved compared with those in column A (on the line above the gaps between crystal blocks): The FWHM (full width at half of maximum) values of the spatial resolutions were 1.6 mm for column B and 2.1 mm for column A.

Moreover, we evaluated the dependence of spatial resolution on the on-focal plane and on the off-focal plane for different source distances from the mid-plane between the two detector heads. A $^{22}$Na point source was placed at various distances (0–10 mm) from the mid-plane to measure the spatial resolution on the mid-plane

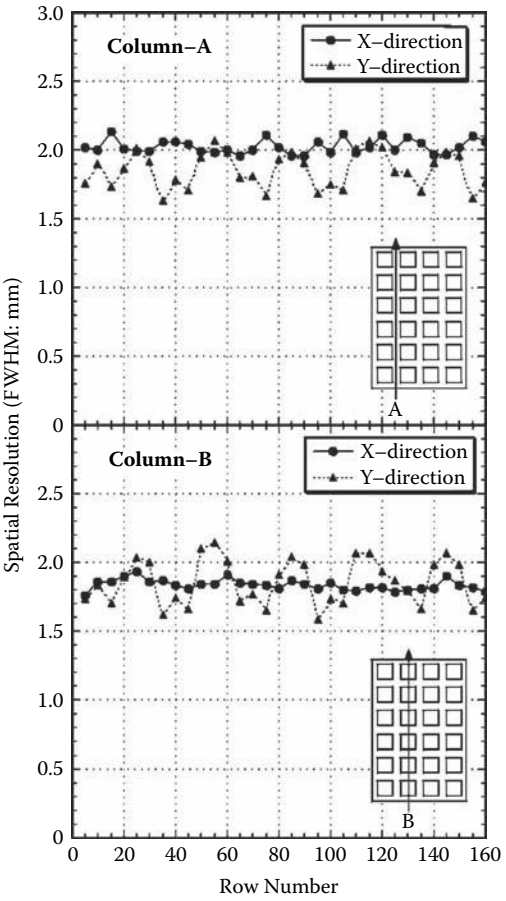



**FIGURE 4.6** Variations in the spatial resolutions along two column lines of the detector head at a D-D distance of 30 cm. (© 2004 by Elsevier B.V. With permission.)

and that on the on-focal plane (the plane on which the source exists) for each source distance. Figure 4.7 shows the changes in the spatial resolutions versus the source distances. The spatial resolution for both the on-focal and off-focal planes deteriorated when sources were located at distances >6 mm from the mid-plane.

#### 4.2.3.2 Sensitivity and Uniformity

The system sensitivity and uniformity on the mid-plane were measured using a flat uniform phantom filled with $^{18}F^{-}$ solution in the narrow mode. The flat phantom, which has inner dimensions of 3-mm thickness × 185-mm width × 250-mm height, contains a radioactive solution and is made of three acrylic plates (each 3 mm thick). The system sensitivity was found to be 107 cps (counts per second) per kBq/mL in the narrow mode. Figure 4.8 demonstrates the focal-plane image measured by the flat phantom filled with $^{18}F^{-}$ solution (so-called flood image) and the profiles, where the image is corrected for the density of coincidence lines back-projected onto the

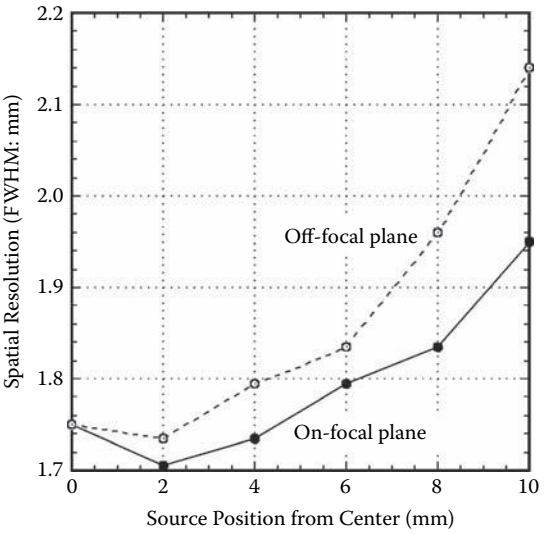



**FIGURE 4.7** Dependence of the spatial resolution on the source position between detectors. Off-focal plane: spatial resolution on the mid-plane. On-focal plane: spatial resolution on the focal planes on which the source is present. (© 2004 by Elsevier B.V. With permission.)

focal plane. The crosshatch pattern on the flood image is caused by the low sensitivities registered at the peripheral regions of detector units. The sensitivity pattern measured for the flat phantom was used to correct the focal-plane images of the other objects to be measured.

The sensitivity can be improved by decreasing the D-D distance or by increasing the acceptance angle for the coincidence lines. On the other hand, the spatial resolution and its uniformity deteriorate by increasing the sensitivity, as mentioned in Section 4.2.3.1. In some measurements for objects having very low-activity tracers, it may be preferable to increase the sensitivity at the cost of spatial resolution.

#### 4.2.3.3 Count-Rate Performance

Figure 4.9 demonstrates the count-rate performance of the system measured using the flat phantom filled with $^{18}F^{-}$ solution, where the effective rate is obtained by subtracting the accidental coincidence rate from the measured one. The maximum effective coincidence count rate was about 20 kcps at an activity concentration of 500 kBq/mL for a 3-mm-thick plane in the narrow mode. If the activity concentration exceeds this condition, the coincidence count-rate increment against the concentration saturates, leading to count loss. The count loss at this high-activity region is corrected by comparing the measured count rate with the ideal count rate, as extrapolated from the measured rates at low activities.

### 4.2.4 Planar Imaging Applicability

We demonstrate here by presenting examples of a rat image observed using PPIS. An anesthetized rat was fixed on an acrylic plate and placed on the mid-plane between

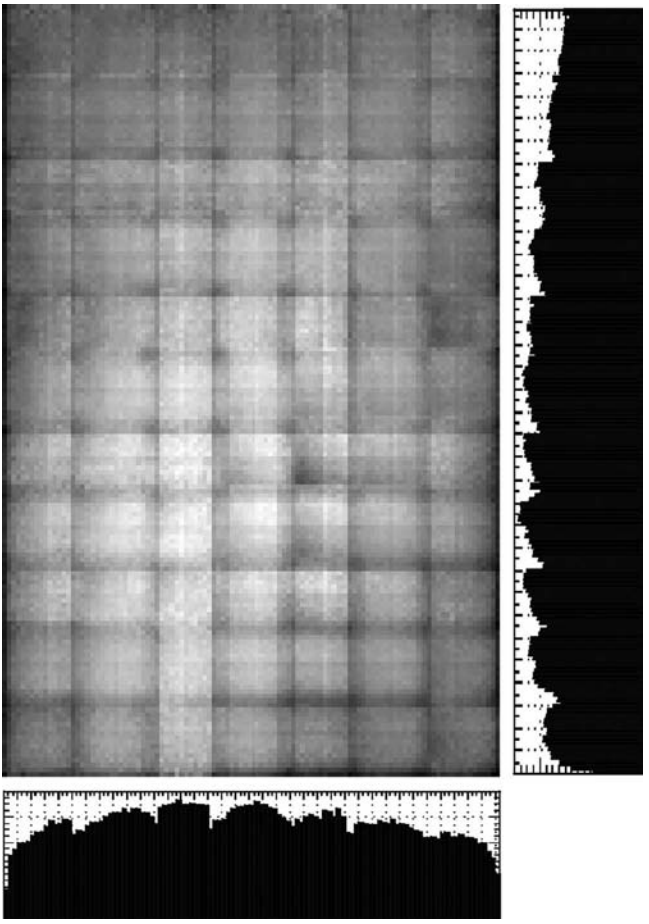


**FIGURE 4.8** Flood image and profiles. Flood images were measured with a flat phantom filled with $^{18}F^{-}$ solution. The bottom and side figure are the profiles for the mid-column and the mid-row, respectively. (© 2004 by Elsevier B.V. With permission.)

the two opposing detectors arranged in a horizontal mode (Figure 4.3A). [$^{18}$F]fluorodeoxyglucose (FDG), a short-lived F-18 isotope ($t_{1/2}$ = 110 min, 8.3 MBq/300 g) was injected intravenously into a rat from the tail vein. The data were acquired with a 10-sec frame interval for the first 10 min after the FDG injection, with a 1-min frame interval for the next 20 min, and with a 10-min frame interval for the last 60 min. Three different frame time images are presented in Figure 4.10. The 10-sec frame images in the upper row demonstrate the capability of this system for a very short period of data acquisition: in the first frame, the introduction of radioactivity from the tail vein was visualized. The subsequent distributions could be clearly seen to be transferred from the liver toward the bowel and then toward the bladder via the kidneys. Thus, a clear image of the radioactivity distribution could be seen from the beginning of the study and the dynamic changes in the activity distribution in a rat

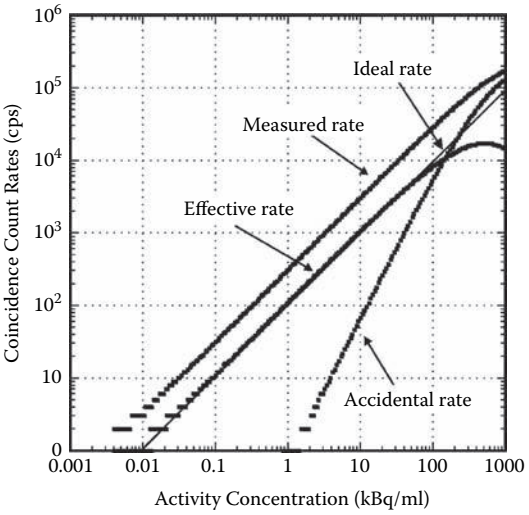



**FIGURE 4.9** Count-rate performance of the system. (© 2004 by Elsevier B.V. With permission.)

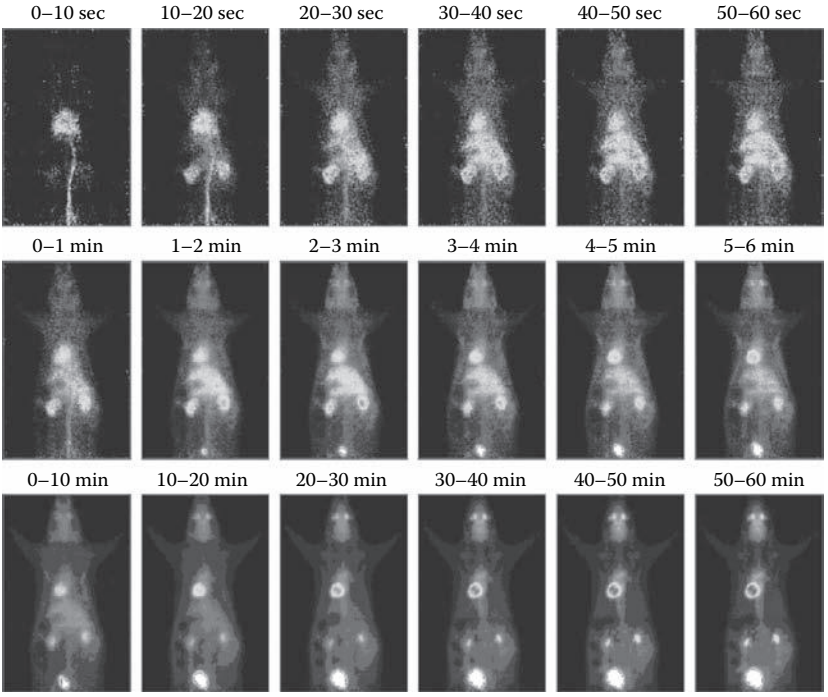



**FIGURE 4.10** Dynamic change of [$^{18}$F] FDG distribution in a rat. Each image in the upper row was measured with a 10-sec frame interval for the first 10 min after the FDG injection. The 1-min frame images for the next 20 min and the 10-min frame images for the last 60 min were shown in the middle row and the lower row, respectively.

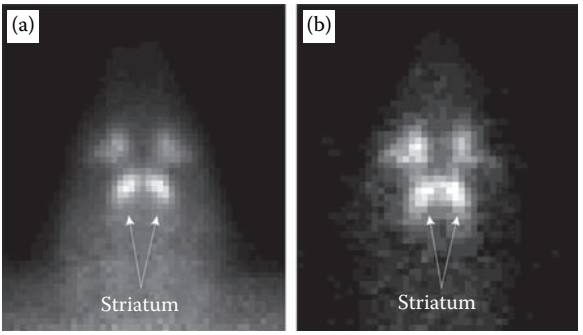



**FIGURE 4.11** Rat brain images by (A) the planar imaging system and (B) the PET system. The images were obtained after the i.v. injection of a brain receptor agent [$^{11}$C]SHC23390. This drug accumulated in the striatum, as can be seen in both images taken after accumulating for the last 20 min in the 60-min scan. (Dose administered: 88.6 MBq/kg [a] versus 136.2 MBq/kg [b]). The radioactivity accumulated in the upper area over the striatum might be that of the intraorbital exocrine gland or lacrimal gland.

model were well measured. The radioactivity injected in this animal was sufficiently low (27.66 MBq/kg) compared with two to five times higher radioactivity required in conventional animal PET devices [25].

Finally, the planar image obtained with the present system was compared with a transaxial image measured by a conventional PET system (SHR7700) [26]. This comparison was performed for rat brain images using a brain-receptor imaging drug, [$^{11}$C]SCH23390, and the results are shown in Figure 4.11. The rat brain images by the planar and PET systems were obtained by accumulation from 41 to 60 min after the administration under the injected dose conditions of 88.6 MBq/kg and 136.2 MBq/kg, respectively. The discrepancy between the outline of both images shows that the planar image suffers from activities existing outside the focal plane as blurred distribution. However, as seen in Figure 4.11, the new device provides almost comparable image quality with PET images, suggesting that PPIS can provide better images for lower radioactivity and for a shorter period of imaging time.

### 4.2.5 Conclusion

Our compact planar imaging system provided superior quality of focal-plane images. The performance of the new planar device was evaluated through experiments using a point source of $^{22}$Na and a flat phantom. The appropriate performance of the new device for the in vivo imaging of small animals was also confirmed by the dynamic measurements of rats. The FOV of the system was large enough to measure the whole body of a rat without scanning. The system offers high-quality images even over a short time frame of measurement. Thus, there is great potential for this new planar imaging system in the high-throughput screening of drugs, as well as for pharmakinetic studies of new drugs with small animals. Moreover, the simple design of the proposed device allows for very versatile use by simply changing the detector posi-

tion. The compactness and low price of this system are expected to enable its use in a wide range of institutions from drug companies to research laboratories.

Finally, the planar image includes information on activities existing outside the focal plane as a blurred distribution, which is different from transaxial images obtained from computed tomography. The planar system, however, is thought to have enough advantages to make up for this weakness.

## 4.3 SMALL-ANIMAL PET SCANNER

### 4.3.1 Introduction

PET dedicated to small animals is one of the powerful tools for molecular imaging, which is required to have high spatial resolution and high sensitivity, resolving the fairly low radiotracer concentration within the target in a small animal. The scanner with the larger solid angle is more suitable for achieving the desired higher sensitivity. However, there is a trade-off between getting the larger solid angle and increasing parallax errors.

The depth of interaction (DOI) information gives the solution for this trade-off. Several ingenious schemes have been proposed to determine the DOI of gamma rays in the scintillator crystals. One of these techniques involves the use of a phoswich detector, which can evaluate the DOI by the use of different fluorescence decay times in scintillators [27, 28]. Another method is to control the light sharing between scintillation crystals [29, 30]. DOI information can also be extracted by using the signal ratio from scintillator crystals sandwiched between two detectors [31]. These schemes require additional special electronics or structural complexity. Thus, we proposed a simple detector structure for DOI measurement. The proposed detector consists of a position-sensitive photomultiplier tube (PS-PMT) and a double-layer array of crystals [32, 33]. This method has also been applied to positron emission mammography [34, 35].

In the following discussion, we introduce a newly developed small-animal PET scanner (Hamamatsu SHR-41000) using the previously mentioned DOI detector to demonstrate its performance.

### 4.3.2 System Construction and Data Processing

The small-animal PET scanner also works similarly with PPIS; the difference is in its detector configuration and reconstruction procedure. In the PET scanner, detector modules are aligned to form a cylindrical shape that surrounds a subject. Here, the term *axial direction* denotes the direction on the surface of the detector module corresponding to the axis of the cylindrical configuration, while the term *transverse direction* denotes the direction corresponding to the circumference of the base of the cylinder.

#### 4.3.2.1 Detector Module

A detector block of the small-animal PET scanner is equipped with a double-layer array of lutetium-yttrium oxyorthosilicate (LYSO) crystals (provided by Photonic Materials, Ltd.), which are coupled to three flat-panel PS-PMTs [36] (Hamamatsu R8400-

00-M256) through a 1.0-mm-thick glass plate. By using the glass plate as a light guide from the LYSO crystal array to the PS-PMTs [37], we can remove crystal gaps between the PMTs to achieve uniform spatial sampling in the axial direction. To provide the DOI detection capability, 32 (transaxially) × 53 (axially) crystal elements are optically coupled to 32 × 54 crystal elements with a shift of half the element pitch in the axial direction of the LYSO crystal array. The dimensions of each crystal element are 1.275 mm (transaxially) × 2.675 mm (axially) × 7 mm (in depth). Since the centroid of the output light distribution for each crystal element is located at a different position, the crystal element can be identified on the image plane derived from the centroid calculation.

The flat-panel PS-PMT has 12 stages of metal-channel dynode and 16 × 16 multiple anodes in a 52.0-mm$^2$ × 14.7-mm-high metal-can package. The effective area of the bialkali photocathode is 49.0-mm$^2$, which provides an approximately useful area ratio of 90%. Since there are still dead spaces between PMTs, a 1-mm-thick glass light guide is placed between the crystal elements and the three PS-PMTs to eliminate the dead area in the axial direction. The detector module—which consists of DOI detectors, voltage dividers, and preamplifier boards—is assembled inside a lightproof case, as shown in Figure 4.12.

Practically, the output signals from multiple anodes of the PS-PMTs distribute nonuniformly, even when detector modules are uniformly illuminated, because there is a nonuniform variation of sensitivity for each anode channel and because the crystal elements generate nonuniform scintillation photons. To compensate for the nonuniform outputs from multiple anodes of the PS-PMT, application-specific integrated circuits (ASICs) having 64-channel variable-gain amplifiers and summing

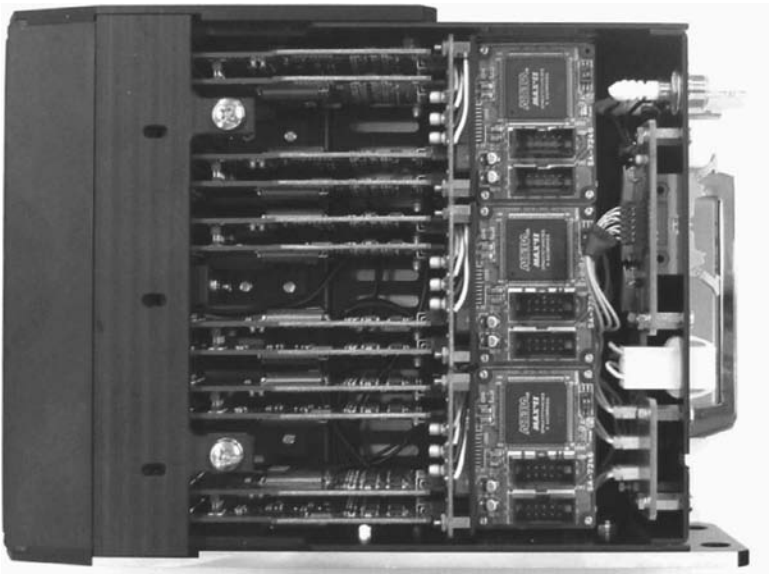


**FIGURE 4.12** Photograph of a detector module consisting of DOI detectors, voltage dividers, and preamplifier boards assembled inside a lightproof case. (© 2008 by IEEE. With permission.)

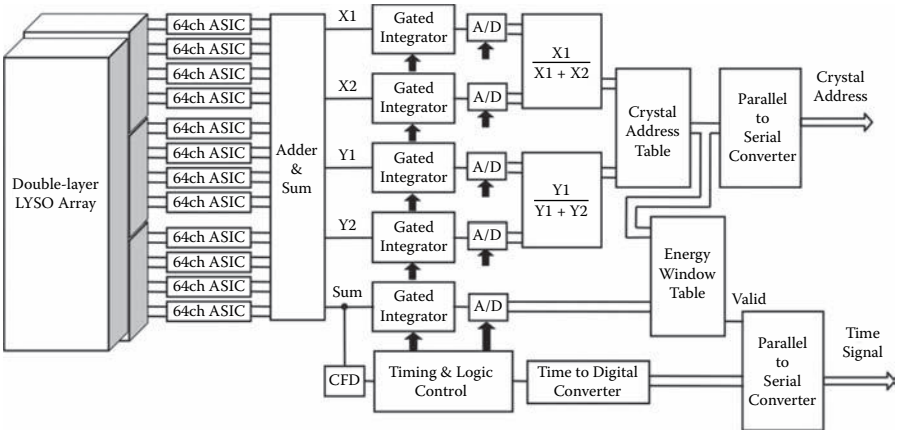



**FIGURE 4.13** Block diagram of the detector-module readout system exhibits how 64-channel ASIC is applied to adjust the gain for various 256-channel output from PMT.

amplifiers are adopted to the front-end circuits. The 256-channel anode outputs from a PS-PMT are fed into four ASICs, in which each anode output is tuned through the variable-gain amplifiers. Figure 4.13 shows the construction of the readout circuit along with a schematic drawing of the detector module.

The position signals and the timing/energy signals from the 12 ASICs in the module are all combined to form position signals ($[X_1, X_2]$, $[Y_1, Y_2]$) and a timing/energy signal from the detector module. In the case of a typical detector module, the energy resolution varies for each crystal element from 19% to 45% with a mean of 25%, where 90% of the crystal elements have energy resolution better than 30%. FWHM of the coincidence timing resolution is measured typically as 2.6 ns using a reference $BaF_2$ detector composed of a 20-mm-diameter × 20-mm crystal and a fast PMT (Hamamatsu R1668) under the choice of the energy window as 350–750 keV.

#### 4.3.2.2 Detector Ring Structure and Gantry

In the present small-animal PET system, 12 DOI detector modules are positioned on a 182-mm-diameter ring to form 107 detector rings with a 1.4-mm pitch. The total crystal number is 41,088 (384 per ring), as shown in the detector ring (Figure 4.14). The FOV forms a cylindrical regime of 100 mm in diameter and 151 mm in the axial length, which is sufficient for covering the whole body of the mouse. The computer-controlled bed for animal subjects is vertically and horizontally adjustable, with an accuracy of 0.1 mm. To obtain higher sampling density and to fill the gaps between the detector modules, the detector rings can be rotated between 0 and 30 degrees, where the angle-adjustable range is equal to the angle between the adjacent modules in the circumference direction of the detector ring. The minimum period of the rotation scan is 10 s, which is sufficiently short for most dynamic studies with small animals. The data for the normalization of the detector sensitivity are obtained by helically scanning a 10-MBq $^{68}Ge$–$^{68}Ga$ rod source having dimensions of 3-mm diameter and 2-mm length. The rod source is shielded to protect the detector modules from unnecessary irradiation. The shield is geometrically designed to provide only

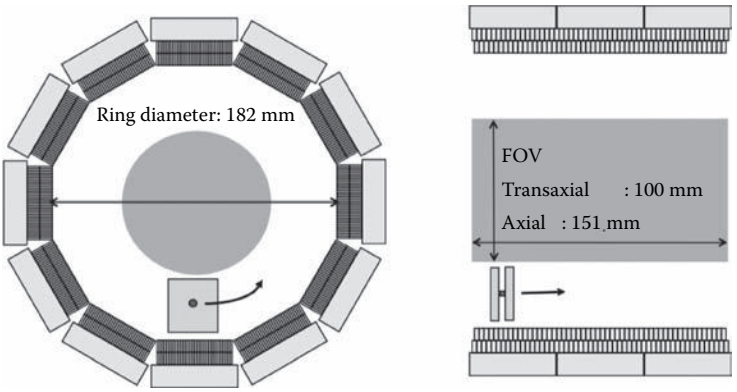



**FIGURE 4.14** Schematic picture of the detector ring exhibits the circular structure of 12 modules and provides the geometrical information of rotating transmission source.


**FIGURE 4.15** Photograph of a small-animal PET scanner.

the coincidence lines inside the FOV. Figure 4.15 shows a photograph of the gantry, while Table 4.1 lists major specifications of the system construction.

#### 4.3.2.3 Electronics

Four position output signals ($X_1$, $X_2$, $Y_1$, $Y_2$) and an energy/timing signal from the detector module are fed into a position analyzer circuit in the gantry, through coaxial

**TABLE 4.1**
**Major Aspects of a Small Animal PET Scanner (Hamamatsu SHR-41000)**

| | |
|---|---|
| **Detector** | |
| Crystal material | LYSO |
| Crystal size | 1.275 mm × 2.675 mm × 7 mm |
| Crystal pitch | 1.4 mm |
| PMT type | flat panel PS-PMT (Hamamatsu R8400-00-M256) |
| Number of crystals | 41,088 (384/ring) |
| Number of PMTs | 36 |
| **Ring geometry** | |
| Number of rings | 107 |
| Ring diameter | 182 mm |
| Ring pitch | 1.4 mm |
| Transaxial FOV | 100 mm |
| Axial FOV | 151 mm |
| **Gantry** | |
| Detector rotation | small angle rotation (0–30 deg.) |
| Positioning | laser projector |
| Source | spiral orbiting $^{68}$Ge-$^{68}$Ga rod source |
| Motion of bed | 300 mm horizontal<br>50 mm vertical |

cables. The energy/timing signal is separated into an energy signal $E$ and a timing signal $T$ in the position analyzer. The position output signals and the energy signal $E$ are integrated over a 200-ns period and digitized with 10-bit ADCs and an 8-bit ADC, respectively. Finally, the positions $x$ and $y$ are calculated from those outputs according to the following relations:

$$x = X_1/(X_1 + X_2)$$

$$y = Y_1/(Y_1 + Y_2)$$

The crystal coding for 3,424 crystals in the detector block is performed by LUT in flash memory. The energy window setting for each crystal is performed using an individually defined LUT. The timing signal $T$ is fed into a constant-fraction discriminator (CFD) that generates a time-mark signal indicating the arrival of a gamma ray. The time-mark signal is used to generate the digitized event time in reference to the system master clock, with 2-ns resolution. To adjust the timing of the signal from each detector module, a programmable delay line having 500-ps time resolution was implemented for each detector module. Each data set consists of a crystal address and time stamp data from 12 detector modules. Any of the 7 opposite

detector modules are treated as detector pairs to detect coincident signals. The width of the coincidence time window can be changed electronically from 6 ns to 18 ns, with 4-ns steps, and is set to 10 ns for LYSO detectors.

#### 4.3.2.4 Data Processing and Image Reconstruction

The coincidence data for all axial acceptance angles are transferred to a data acquisition unit consisting of a commodity personal computer operated with LINUX, where the coincidence data are stored in "list mode." The list mode data are binned into the sinograms defined by the axial coordinate and the ring difference, while taking account of the DOI information. For two-dimensional (2-D) reconstruction, the list mode data are combined to a three-dimensional (3-D) data set, and the 3-D data set is converted to 213 slices of 2-D sinograms through a Fourier rebinning algorithm (FORE) [38], restricting maximum ring difference as $\delta = \pm 25$ with the compression of span 3, or $\delta = \pm 42$ with that of span 5. The axial sampling distance is 0.7 mm. Images are reconstructed from the 2-D sinograms with standard 2-D filtered back projection (FBP) or a fast iterative-reconstruction algorithm, 2-D DRAMA (Dynamic RAMLA) [39]. For the 3-D direct reconstruction, 3-D DRAMA and subsetized list mode EM (SLEM) [40] are adopted into the system.

To normalize the detector sensitivities, a component-based normalization (CBN) method is used [41]. The attenuation and scatter corrections are not yet implemented.

### 4.3.3 Performance Evaluation

The evaluation of the PET system performance has been standardized for a whole-body PET by NEMA NU2-2001. However, since the NEMA test procedure cannot be applied to small-animal PET scanners, we prepared small-sized phantoms for a rat for evaluating the present small-animal PET system. The procedures, such as data processing and analysis, mostly followed the NEMA protocol. The experiments described below were performed with a 10-ns coincidence time window width, setting the energy window to 350–750 keV.

#### 4.3.3.1 Detector-Pair Resolution

We evaluated the limitation of spatial resolution by determining, from the hardware specifications of the detector module, the detector-pair resolution. The dimensions of the crystal elements located on the detector rings are 1.275 mm in the transverse direction (i.e., the direction of circumference of the detector ring) and 2.675 mm in the axial direction. The intrinsic spatial resolutions of the detector pair were measured in the two orthogonal directions. The measurements were performed by scanning a 1.0-mm-diameter $^{22}$Na point source from the center of the FOV (CFOV) with 0.2-mm steps in the transverse and axial directions. The coincidence count rate was plotted as a function of the source position and was fitted with a Gaussian function, whose FWHM value gives the detector-pair resolution. Figures 4.16(a) and 4.16(b) show the response curves in the transverse and axial directions, respectively.

The measured FWHM values in the transverse direction ranged from 1.75 to 1.90 mm, with an average of 1.83 mm (Figure 4.16a). On the other hand, the FWHM

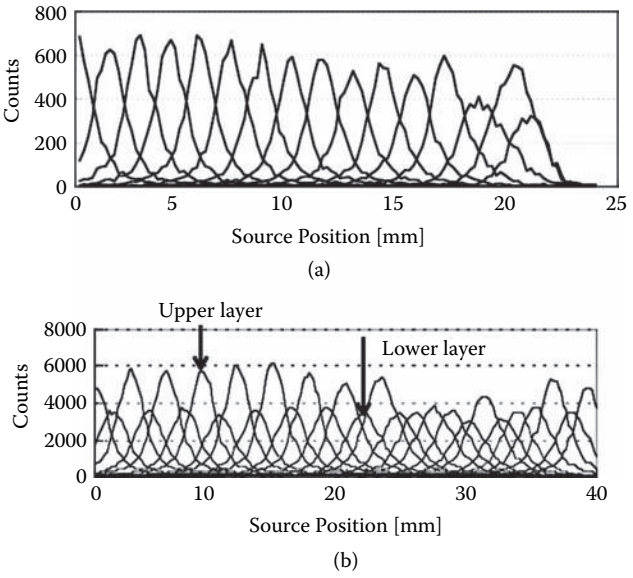



**FIGURE 4.16** Intrinsic spatial resolution for (a) transverse and (b) axial directions. (© 2008 by IEEE. With permission.)

values in the axial direction revealed the systematic difference for signals originated from upper- and lower-layer crystals: the FWHM for the upper layer ranged from 2.82 to 2.86 mm with an average of 2.84 mm, whereas that for the lower layers ranged from 3.19 to 3.35 mm with an average of 3.26 mm. Additionally, the axial detector-pair resolutions for the crystals on the PMT gaps in the axial direction were slightly poorer than those for the crystals right on the PMTs. The total FWHM values involving the effects of layer dependence and PMT gaps ranged from 2.82 to 4.08 mm with an average of 3.37 mm.

#### 4.3.3.2 Spatial Resolution of the System

We describe here the system's total resolution, including the previously mentioned detector-pair resolution and the effects of reconstruction. Note that the term *transaxial resolution* denotes the spatial resolution evaluated from the cross-sectional images.

To evaluate the transaxial resolution, we prepared a line source by filling a glass capillary (0.3-mm inner diameter) with $^{18}F^{-}$ solution. The transaxial resolution was measured with the line source placed at 0, 5, 10, 20, 30, and 40 mm from the center along the central axis of the gantry hole. The data were acquired with and without detector rotation for comparison. The scan period was 5 min at each position to collect events of more than 100k counts. The list mode data were rebinned into 2-D sinograms by FORE with $\delta = \pm 25$ and reconstructed by FBP with a ramp filter. For the measurement of the spatial resolution, the image was reconstructed as a $256 \times 256$ matrix with a bin width of 0.4 mm, and the FWHM values of the profiles were determined in the directions perpendicular to the line source using linear interpolation. The FWHM values were calculated for 11 adjacent direct planes and averaged to

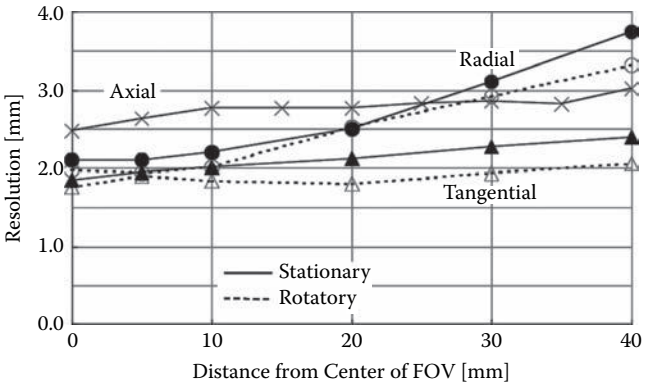



**FIGURE 4.17** Spatial resolution measured from the reconstructed image.

obtain the mean resolution value. The axial resolution was measured with a 40-mm-long line source filled with $^{18}F^-$ solution placed radially at the CFOV. The source was then moved along the axis of the plane in steps of 0.175 mm. Figure 4.17 shows the resolution (FWHM) in the radial and tangential directions with and without detector rotation scan, as well as that in the axial direction. The transaxial resolution at the CFOV is less than 2.0 mm, and the axial resolution is around 2.8 mm in the whole FOV. The resolution in both the radial and tangential direction is improved by the detector rotation scan.

#### 4.3.3.3 Sensitivity

The source for the sensitivity measurement was a glass capillary (1.0 mm in diameter × 2.0 mm in length) filled with 0.47-MBq $^{18}F^-$ solution. The activity was sufficiently low so that the dead-time loss was negligible. The source was located at the CFOV, and the coincidence counts for all axial acceptance angles were collected. The absolute system sensitivity was 8.1% at the CFOV owing to the large axial FOV.

#### 4.3.3.4 Scatter Fraction and Noise-Equivalent Count Rate

The scatter fraction and the count-rate performance were measured for a rat-sized solid acrylic cylindrical phantom whose diameter and overall length were 6 and 15 cm, respectively. The rat-sized phantom had a 2-mm-diameter hole drilled parallel to the central axis of the cylinder at a radial distance of 1.5 cm. A glass capillary (1.0-mm inner diameter and 15-cm long) filled with $^{18}F^-$ solution was inserted into the hole. The phantom, whose initial activity was 17.0 MBq, was located at the center of both transaxial and axial FOV in the scanner using the animal bed. The data were acquired in list mode to generate sinograms for each acquisition frame and for each slice. The oblique sinograms were collapsed into a single sinogram by single-slice rebinning. The coincidence background noise due to the intrinsic radioactivity of $^{176}Lu$ contained in LYSO was subtracted from the measured coincidence events. The count rate of the coincidence background noise was about 1.5 kcps and was independent of the activity in the phantom.

The scatter fraction (SF) was derived using the final acquisition data to give an averaged SF of 34.6% for the phantom. The noise-equivalent count rate (NECR) was calculated from the true coincidence ($T$), random coincidence ($R$) measured by delayed coincidence technique, and scatter coincidence ($S$) as follows:

$$\text{NECR} = T^2/(T + S + 2fR)$$

The constant, $f$, is the fraction of a projection subtended by the phantom [42], and $f = 0.6$ for this phantom. Figure 4.18 shows the true count rate, random count rate, and NECR curves. The peak count rate of NECR was 72.6 kcps at 30.7 kBq/mL.

#### 4.3.3.5 In Vivo Imaging Studies of Animals

To demonstrate the image performance of the scanner, a Sprague-Dawley (SD) rat (8 weeks, 266 g) was measured with the tracer of $^{18}F^-$. The images were reconstructed using both FORE + FBP and SLEM with finite-resolution modeling, where no attenuation and scatter corrections were employed.

Figure 4.19 shows the maximum intensity projection (MIP) views of bone images of the rat, measured at three bed positions. Data acquisition was started 72 min after the injection of 8.0 MBq of $^{18}F^-$ and continued during 20-min duration at each bed position. In SLEM, finite-resolution effects (such as positron range, photon noncollinearity, and intrinsic detector resolution) can be modeled by shift-invariant kernels; hence SLEM with finite-resolution modeling gives sharper images than other methods.

### 4.3.4 Discussion

We introduced a newly developed small-animal PET scanner (Hamamatsu SHR-41000) using DOI detectors with large detection areas and evaluated its initial

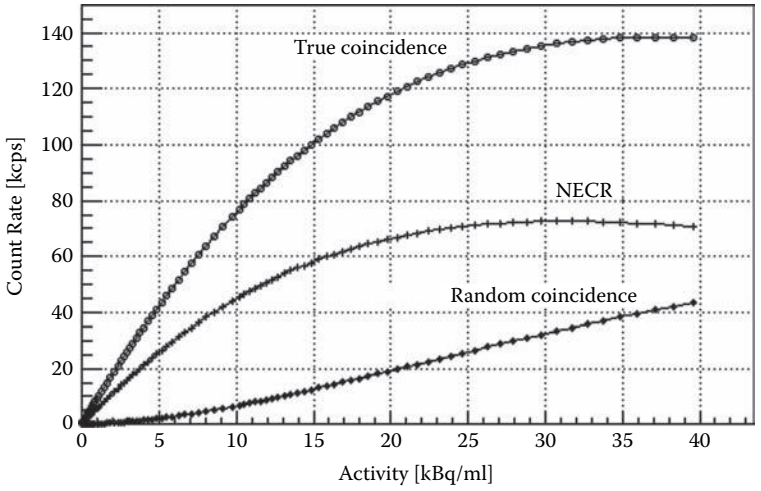



**FIGURE 4.18** Count-rate capability as a function of the line-source activity divided by the total volume of the cylindrical rat-sized phantom (424 mL). (© 2008 by IEEE. With permission.)

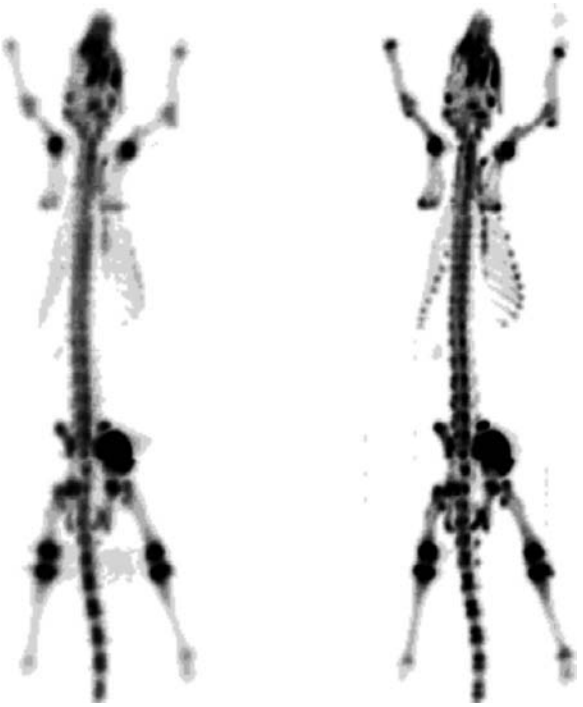


**FIGURE 4.19** The bone image of a whole body of a 256-g rat. The left image was reconstructed using FORE + FBP, and the right image was reconstructed using SLEM. (© 2008 by IEEE. With permission.)

performance. The large axial FOV of 151 mm sufficiently covers the whole body of a mouse and can measure a rat with two or three steps of the bed. Because the present PET scanner has higher sensitivity owing to its larger solid angle, it is possible to perform animal experiments with small doses of radiotracers, as the NECR characteristic shows. On the other hand, the maximum NECR is limited to 72.6 kcps for the rat-sized phantom at the cost of the large detection area of the detectors.

The spatial resolution of 2 mm at the CFOV is larger than that expected from the crystal size. This is caused by the incompleteness of crystal separations brought from current preamplifiers in the system. The current ASICs still have some electronic noise. Of the outputs from the 12 ASICs, 48 are used for the position calculations, while the signal information is mainly included in only 4 outputs. The capability to identify each crystal segment is degraded by summing the electronic noise from the ASICs.

One of the benefits expected with the use of the DOI detector is to suppress the degradation of the spatial resolution in the radial direction. We compared the resolutions with and without the use of DOI information, but there is no difference, not even in the periphery. This is also caused by the event crosstalk between the upper and lower crystal segments, again due to the incompleteness of crystal separations.

Additionally, the light spread caused by the propagation via the light guide between the crystal array and PMTs can also degrade the crystal separations. We have redesigned the circuit to utilize the signals more efficiently, for example, by using threshold amplifiers [43]. Figure 4.20 shows the flood images of the detector block under uniform irradiation of 511-keV gamma rays. We compared the use of the current preamplifiers and the redesigned preamplifiers including new ASICs with threshold amplifiers. The crystal separations are significantly improved with the use of the threshold amplifiers. Moreover, the preamplifier circuits were redesigned to increase the high-voltage supply for PMTs for the purpose of improving the signal-to-noise ratio. We are presently attempting to improve the threshold amplifier function of the new ASICs, with the expectation of improving the DOI measurement capability and the spatial resolution.

The initial evaluation results and demonstrated images for animal studies indicate that the new small-animal PET scanner should prove to be a useful device for in vivo molecular imaging. There is still much room for improvement in both the system design and the software; nevertheless, better performance is expected in the future.

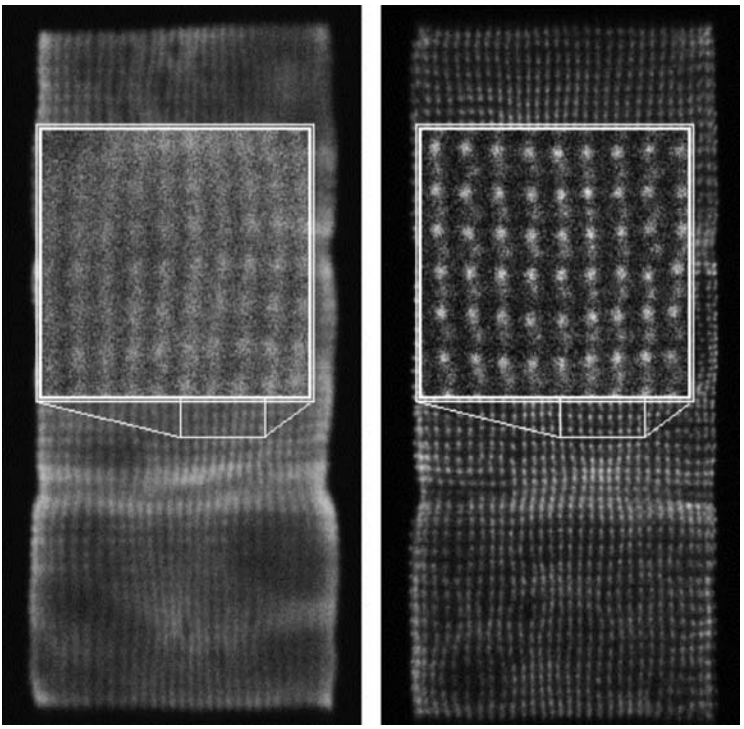


**FIGURE 4.20** The flood images of the detector block. The left image was obtained with the current preamplifiers, and the right image was obtained with the redesigned preamplifiers. (© 2008 by IEEE. With permission.)

## 4.4 SUMMARY

In this chapter, we have described in detail the system construction and fundamental specifications of the planar positron imaging system (PPIS) and the small-animal PET scanner. While the systems and the detectors continue to improve, both systems have already been adopted in practical use. In the case of the PPIS installed at the PET Center of Hamamatsu Photonics K.K., a feasibility study of quantitative imaging has been performed for further system evaluation [44]. The PPIS has also been used to study the use of radiotracers as tumor-imaging agents [45]. In the case of the small-animal PET installed at the Biomedical Imaging Research Center of Fukui University, a feasibility study of probes for in vivo monitoring of adenoviral-mediated gene therapy has been conducted [46]. Molecular imaging using both systems has been studied independently and complementarily. Each system has its own advantages and disadvantages, which suggests that their combined use would be much more effective in developing new drugs.

## REFERENCES

1. Townsend, D. W. 2008. Multimodality imaging of structure and function. *Phys. Med. Biol*. 53:R1–R39.
2. Rudin, M., and R. Weissleder. 2003. Molecular imaging in drug discovery and development. *Nat. Rev. Drug Disc.* 2:123–131.
3. Cherry, S. R. 2004. In vivo molecular and genomic imaging: New challenges for imaging physics. *Phys. Med. Biol*. 49:R13–R48.
4. Uchida, H., T. Okamoto, T. Ohmura, K. Shimizu, N. Satoh, T. Koike, and T. Yamashita. 2004. A compact planar positron imaging system. *Nucl. Instrum. Methods Phys. Res. A* 516:564–574.
5. Yamada, R., M. Watanabe, T. Omura, N. Sato, K. Shimizu, M. Takahashi, K. Ote, A. Katabe, T. Moriya, K. Sakai, T. Yamashita, and E. Tanaka. 2008. Development of a small animal PET scanner using DOI detectors. *IEEE Trans. Nucl. Sci*. 55:906–911.
6. Brownell, G. L., C. A. Chesler, D. Correia, et al. 1977. Transverse section imaging of radionuclide distribution in the heart, lung and brain. In *Reconstruction tomography in diagnostic radiology and nuclear medicine*, ed. M. M. Ter-Pogossian, M. E. Phelps, and G. L. Brownell, 293–307. Baltimore: University Park Press.
7. Budinger, T. F., S. E. Derenzo, G. T. Gullberg, W. L. Greenberg, and R. H. Huesman. 1977. Emission computer assisted tomography with single-photon and positron annihilation photon emitters. *J. Comput. Assist. Tomogr*. 1:131–145.
8. McKee, B. T. A., H. B. Dinsdale, M. J. Hogan, D. C. N. Howse, J. Kulick, H. B. Mak, and T. T. Stewart. 1988. Description and performance of a prototype PET system for small volume imaging. *Nucl. Inst. and Meth*. 269:392.
9. Muehllehner, G., M. Geagan, P. Countryman, and P. Nellemann. 1995. SPECT scanner with PET coincidence capability. *J. Nucl. Med*. 36:70P.
10. Patton, J. A., Y. Hefetz, M. D. Shone, M. B. Stoner, and M. P. Sandler. 1997. Measured coincidence imaging parameters of a clinical dual head scintillation camera. *J. Nucl. Med*. 38:221P.
11. Fryer, T. D., R. W. Barber, N. J. Bird, et al. 1998. PET imaging with the Picker gamma camera PCD system. *Br. J. Radiol*. 71:15.

12. Brasse, D., M. Tararine, O. Lamar, and B. Bendriem. 1997. Investigation of noise equivalent count rate in positron imaging using a dual head gamma camera. In *Conf. Record IEEE Nucl. Sci. Symp. Conf. Rec*. Vol. 2, 1223–1227.
13. S.D.Wollenweber, B.M.W. Tsui, D. S. Lalush, E. C.Frey, W.H. McCartney, C. Tocharoenchai, T.D. Rempe, and V. Simcic, Initial characterization of the Siemens E.CAM+: a dual-detector camera with coincidence imaging capability, *Conf. Record IEEE Nuclear Science Symp.*, vol. 3, Tronto, Ont, Nov. 1998, pp. 1470–1476.
14. Lewellen, T. K., R. S. Miyaoka, and W. L. Swan. 1999. PET imaging using dual-headed gamma cameras: An update. *Nucl. Med. Commun*. 20:5–12.
15. N. Keutgen, S. Matsuhashi, C. Mizuniwa, T. Ito, T. Fujimura, N.S. Ishioka, S. Watanabe, T. Sekine, H. Uchida, and S. Hashimoto, Transfer function analysis of positron-emitting tracer imaging system (PETIS) data, *Appl. Radiat. Isotopes*, vol. 157, pp. 225–233, 2002.
16. H. Nakanishi, S. Kiyomiya, T. Tsukamoto, H. Tsukada, H. Uchida, and S.Mori, Water (15OH2) flow in rice is regulated by the concentration of nutrients as monitored by positron multi-probe system (PMPS), *Soil Sci. Plant Nutr.*, vol. 48, pp. 759–762, 2002.
17. H. Fujikake, A. Yamazaki, N. Ohtake, K. Sueyoshi, S. Matsuhashi, T. Ito, C. Mizuniwa, T. Kume, S. Hashimoto, N. S. Ishioka, S. Watanabe, A. Osa, T. Sekine, H. Uchida, A. Tsuji and T. Ohyama, Quick and reversible inhibition of soybean root nodule growth by nitrate involves a decrease in sucrose supply to nodules, *J. Exp. Bot.*, vol. 54(386), pp. 1379–1388, 2003.
18. T. Tsukamoto, H. Uchida, H.Nakanishi, S. Nishiyama, H. Tsukada, S. Matsuhashi, N. K. Nishizawa, and S. Mori, H215O Translocation in Rice Was Enhanced by 10μM 5-Aminolevulinic Acid as Monitored by Positron Emitting Tracer Imaging System (PETIS), *Soil Sci. Plant Nutr.*, vol. 50, pp. 1085–1088, 2004.
19. A.J. Keutgen, N. Keutgen, S. Matsuhashi, C. Mizuniwa, T. Ito, T. Fujimura, N. S. Ishioka, S. Watanabe, A. Osa, T. Sekine, H. Uchida, A Tsuji, and S. Hashimoto, Input-output analysis of in vivo photoassimilate translocation using Positron-Emitting Tracer Imaging System (PETIS) data, *J. Exp. Bot.*, vol. 56, pp. 1419–1425, 2005.
20. S. Matsuhashi, S.Fujimaki, H. Uchida, N. S. Ishioka, and T. Kume, A new visualization technique for the study of the accumulation of photoassimilates in wheat grains using [11C]CO2, *Appl. Radiat. Isotopes*, vol. 64, pp. 435–440, 2006.
21. E. Hattori, H. Uchida, N. Harada, M. Ohta, H. Tsukada, Y. Hara and T. Suzuki, Incorporation and translocation of 2-deoxy-2-[18F]fluoro-d -glucose in Sorghum bicolor (L.) Moench monitored using a planar positron imaging system, *Planta*. vol. 227, pp. 1181–1186, 2008.
22. T. Tsukamoto, H. Nakanishi, H. Uchida, S. Watanabe, S. Matsuhashi, S. Mori, N. Nishizawa, 52Fe Translocation in Barley as Monitored by a Positron-Emitting Tracer Imaging System (PETIS): Evidence for the Direct Translocation of Fe from Roots to Young Leaves via Phloem, *Plant and Cell Physiology.*, vol. 50(1), pp. 48–57, 2009.
23. Hamamatsu Photonics K.K. 2001. Position sensitive photomultiplier tube R8520-00-C12. Preliminary data sheet, cat. no. TPMH1276E01.
24. Uchida, H., T. Yamashita, M. Iida, and S. Muranatsu. 1986. Design of mosaic BGO detector system for positron CT. *IEEE Trans. Nucl. Sci*. 33:464–467.
25. Handa, N., Y. Magata, E. Tadamura, T. Mukai, T. Nishina, S. Miwa, Y. Sakakibara, T. Nomoto, J. Konishi, K. Nishimura, and M. Komeda. 2002. Quantitative fluorine 18 deoxyglucose uptake by myocardial positron emission tomography in rats. *J. Nucl. Cardiol*. 9:616–621.
26. Watanabe, M., H. Okada, K. Shimizu, T. Omura, E. Yoshikawa, T. Kosugi, S. Mori, and T. Yamashita. 1997. A high resolution animal PET scanner using compact PS-PMT detectors. *IEEE Trans. Nucl. Sci*. 44:1277–1282.

27. Casey, M. E., L. Eriksson, M. Schmand, M. S. Andreaco, M. Dahlbom, R. Nutt, and M. Paulus. 1997. Investigation of LSO crystals for high spatial resolution positron emission tomography. *IEEE Trans. Nucl. Sci*. 44:1109–1113.
28. Yamamoto, S., and H. Ishibashi. 1998. A GSO depth of interaction detector for PET. *IEEE Trans. Nucl. Sci*. 45:1078–1082.
29. Murayama, H., I. Ishibashi, H. Uchida, T. Omura, and T. Yamashita. 1998. Depth encoding multicrystal detectors for PET. *IEEE Trans. Nucl. Sci*. 45:1152–1157.
30. Miyaoka, R. S., T. K. Lewellen, H. Yu, and D. L. McDaniel. 1998. Design of a depth of interaction (DOI) PET detector module. *IEEE Trans. Nucl. Sci*. 45:1069–1073.
31. Moses, W. W., and S. E. Derenzo. 1994. Design studies for a PET detector module using a PIN photodiode to measure depth of interaction. *IEEE Trans. Nucl. Sci*. 41:1441–1445.
32. Yamashita, T., M. Watanabe, K. Shimizu, and H. Uchida. 1990. High resolution block detector for PET. *IEEE Trans. Nucl. Sci*. 37:589–593.
33. Liu, H., T. Omura, M. Watanabe, and T. Yamashita. 2001. Development of a depth of interaction detector for γ-rays. *Nucl. Instr. Meth. A* 459:182–190.
34. Thompson, C. J., K. Murthy, Y. Picard, I. N. Weinberg, and R. Mako. 1995. Positron emission mammography (PEM): Promising technique for detecting breast cancer. *IEEE Trans. Nucl. Sci*. 42:1012–1017.
35. Zhang, N., C. J. Thompson, D. Jolly, F. Cayouette, and S. Kecani. 2002. A prototype modular detector design for high resolution positron emission mammography imaging. *IEEE Trans. Nucl. Sci*. 50:1624–1629.
36. Kyushima, H., H. Shimoi, A. Atsumi, M. Ito, K. Oba, and Y. Yoshizawa. 2000. The development of flat panel PMT. In *IEEE Nucl. Sci. Symp. Conf. Rec*. Vol. 1, 7/3–7/7.
37. Rouze, N. C., M. Schmand, S. Siegel, and G. D. Hutchins.2004. Design of a small animal PET imaging system with 1 microliter volume resolution. *IEEE. Trans. Nucl. Sci*. 51:757–763.
38. Defrise, M. 1995. A factorization method for the 3D X-ray transform. *Inverse Problems* 11:983–994.
39. Tanaka, E., and H. Kudo. 2003. Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography. *Phys. Med. Biol*. 48:1405–1422.
40. Rahmim, A., M. Lenox, A. J. Reader, C. Michel, Z. Burbar, T. J. Ruth, and V. Sossi. 2004. Statistical list-mode image reconstruction for the high resolution research tomography. *Phys. Med. Biol*. 49:4239–4258.
41. Badabi, R. D., and P. K. Marsden. 1999. Developments in component-based normalization for 3D PET. *Phys. Med. Biol*. 44:571–594.
42. Strother, S. C., M. E. Casey, and E. J. Hoffman. 1990. Measuring PET scanner sensitivity: Relating count rates to image signal-to-noise ratios using noise equivalent count. *IEEE. Trans. Nucl. Sci*. 37:783–788.
43. Park, B., T. K. Lewellen, R. S. Miyaoka, and P. Kinahan. 2004. Analog processing board for the MiCES small animal PET scanner. In *IEEE Nucl. Sci. Symp. Conf. Rec*. Vol. 6, 3920–3923.
44. Uchida, H., K. Sato, T. Kakiuchi, D. Fukumoto, and H. Tsukada. 2008. Feasibility study of quantitative radioactivity monitoring of tumor tissues inoculated into mice with a planar positron imaging system (PPIS). *Ann. Nucl. Med*. 22:57–64.
45. Tsukada, H., K. Sato, D. Fukumoto, S. Nishiyama, N. Harada, and T. Kakiuti. 2006. Evaluation of D-isomers of O-$^{11}$C-methyl tyrosine and O-$^{18}$F-fluoromethyl tyrosine as tumor-imaging agents in tumor-bearing mice: Comparison with L- and D-$^{11}$C-methionine. *J. Nucl. Med*. 47:689–688.
46. Lohith, T. G., T. Furukawa, T. Mori, M. Kobayashi, and Y. Fujibayashi. 2008. Basic evaluation of FES-hERL PET tracer-reporter gene system for in vivo monitoring of adenoviral-mediated gene therapy. *Mol. Imaging Biol*. 10:245–252.

# 5 PET Front-End Electronics

*Christoph Werner Lerche and*
*Vicente Herrero Bosch*

## CONTENTS



## 5.1 INTRODUCTION

Most of the gamma-ray detectors employed in positron emission tomography (PET) applications provide output signals that need further amplifying and processing to carry out an optimal analog-to-digital conversion. Front-end electronics implement all the necessary analog processing from the detector's output to the input of the data-acquisition and control (DAC) system. A typical front-end architecture usually includes an input preamplifying stage and an output shaping stage.

Preamplifiers collect the detector's output signals, which, due to their reduced amplitude and thus high sensibility to noise, must be carefully amplified as soon as possible. Occasionally, the preamplifying stage is omitted if the signal-to-noise ratio provided by the detector is high enough and there is no need to isolate the detector from the other parts of the front end.

The shaping stage produces the slow-changing signals required by analog-to-digital converters (ADCs), thereby overcoming bandwidth limitations. Obviously,

the desired characteristics of output signals should remain unaltered throughout the shaping process.

Some kinds of detectors, such as position-sensitive photomultiplier tubes (PSPMTs), make use of charge-division circuits between preamplifying and shaping stages, which reduce the number of channels to be converted in DAC. This reduction allows a decrease in the front-end complexity, especially when dealing with the latest high-resolution position-sensitive detectors. Basically, these charge-division circuits make an analog computation to obtain some information related to the gamma event detected, such as its position on the detector's surface or its depth inside the scintillator crystal [1].

Due to differences among the available detector types, the front-end electronics must be custom designed in order to obtain the best performance. An equivalent model of the detector will help throughout the design process.

## 5.2 PREAMPLIFYING STAGE

PET-oriented detectors must work in pulse mode, sometimes called photon counting, to obtain information about energy, timing, etc., of the detected event. The nature of the pulse signal generated by the detector determines the architecture of the preamplifier's input. The first step in the design procedure will be to characterize the detector, which eventually will end up with its electrical equivalent model. Depending on the temporal behavior of the detector-preamplifier system, two different types of scenarios arise:

- Preamplifier just follows the input signal with a high degree of fidelity up to a certain bandwidth.
- Preamplifier integrates the input signal, giving information about the pulse area.

There are two types of detector architectures: direct and indirect. The main difference between them lies in the physical procedures involved. The direct detector changes incident gamma photons into an electrical charge that is captured by the preamplifier. The indirect detector is composed of two blocks: a scintillator component and a visible-light detector. The scintillator absorbs the gamma photons and emits visible-light photons, which are captured by the photodetector. Also, there are several types of photodetectors that basically differ in the amount of electrical charge produced. Photomultipliers produce a high amount of charge, which can be processed as a current magnitude. Other devices such as APD (avalanche photo diode) produce small amounts of charge due to their lower gain. This fact makes necessary an amplification procedure similar to that employed in direct detectors.

### 5.2.1 Equivalent Models of Detectors

In this section, some electrical equivalent models of detectors are introduced. Two different types of photomultipliers are covered: photomultiplier tubes (PMT)

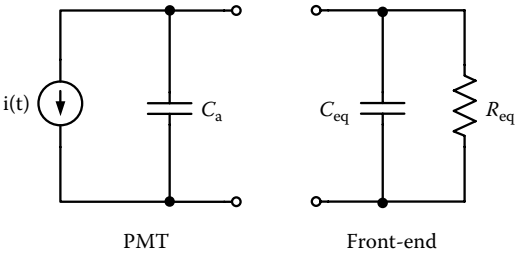



**FIGURE 5.1** Equivalent circuit of PMT connected to front end.

and silicon photomultipliers (SiPM). Also, a general model of direct detectors is explained that can be applied to APD-based indirect detectors, although in the latter case the amount of charge obtained is higher.

#### 5.2.1.1 Photomultiplier Tube Detector Equivalent Circuit

The equivalent circuit of this device is quite simple (Figure 5.1). Current source gives the output pulse signal, and $C_a$ models the equivalent capacitance of the output anode of the PMT. The circuit also applies to PSPMT, where each output needs an individual model. Anode capacitance values are in the range of picofarads to tenths of picofarads. PSPMTs show a wide dispersion in their parasitic capacitances, which can reach 50%, thus affecting temporal response of the device as a function of spatial position.

The current source has to model the behavior of the scintillation crystal and the electron-multiplying process carried out by the PMT [2]. The scintillation light generated by the crystal can be approximately described by Equation (5.1)

$$I = I_0 e^{-\frac{t}{\lambda}} \tag{5.1}$$

This exponential law has a decay constant λ, which depends on crystal composition and a nearly zero rise time. This approximation can be made with no loss of accuracy if the front end's time constant is larger than the light-pulse rise time. Following the signal chain, the PMT can be treated as just an amplifying photosensor that holds true for TTS* lower than the scintillator decay time constant. This means that photons reaching the PMT's surface are converted to a certain number of electrons following the same arriving order. Output current signals are composed by superposition of the individual PMT responses to each visible-light photon. As a consequence, considering a mean photon flow and a fast response of the PMT,† the resulting output current pulse can be described as

$$i(t) = i_0 e^{-\lambda t} \tag{5.2}$$

* Transit, time spread: statistical dispersion of electron transit time in electron-multiplying stage in PMTs.

† Rise times of most PMTs are on the order of 1 to 2 ns.

where $i_0$ is the initial and maximum current that can be obtained from the total amount of charge-generated $Q$ as

$$Q = \int_0^\infty i(t)dt = i_0 \int_0^\infty e^{-\lambda t}dt = \frac{i_0}{\lambda} \tag{5.3}$$

$$i_0 = \lambda Q \tag{5.4}$$

Finally, Equation (5.5) is a fair approximation of an output current pulse obtained from an indirect detector based on a PMT.

$$i(t) = \lambda Q e^{-\lambda t} \tag{5.5}$$

If a PSPMT is used, each output will have a model like Figure 5.1. However, the amount of charge-generated $Q$ will depend on the light distribution on the detector's surface. This light distribution can be estimated [1] as

$$L = \frac{d_{\text{eff}}}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + d_{\text{eff}}^2}} \tag{5.6}$$

where $d_{\text{eff}}$ is the depth of the impact of the gamma ray; $x$, $y$ is the location of the point where we want measure the light; and $x_0$, $y_0$ is the position of impact on the detector's surface.

To develop a study of the influence of the front end on output signal characteristics, the expression of anode output voltage is obtained from Equation (5.5) as

$$i(t) = i_{C_{\text{tot}}} + i_{R_{\text{eq}}} \tag{5.7}$$

where $C_{\text{tot}}$ is the sum of $C_{\text{a}}$ and $C_{\text{eq}}$, which models parasitic capacitance from the front end, cables, connections, etc.

$$i(t) = C_{\text{tot}} \frac{dV(t)}{dt} + \frac{V(t)}{R_{\text{eq}}} \tag{5.8}$$

$$\frac{dV(t)}{dt} + \frac{1}{R_{\text{eq}}C_{\text{tot}}} V(t) = \frac{\lambda Q}{C_{\text{tot}}} e^{-\lambda t} \tag{5.9}$$

Resolving this last differential equation with initial condition $V(0) = 0$, we obtain

$$V(t) = \frac{1}{\lambda - \theta} \cdot \frac{\lambda Q}{C_{\text{tot}}} (e^{-\theta t} - e^{-\lambda t}) \tag{5.10}$$

where $\theta = \frac{1}{R_{\text{eq}}C_{\text{tot}}}$ is the time constant of the whole circuit—detector plus front end. In the following sections, a characterization of the preamplifier behavior will use both $\theta$ and $\lambda$ time constants.

### 5.2.1.2 Silicon Photomultiplier Detector Equivalent Circuit

Semiconductor-based photomultiplier devices show promising features and could become a real alternative to traditional PMTs in the near future [3–5]. Their intrinsic equivalent model is somewhat more complicated than that shown in Figure 5.1 due to the complexity of their structure [6]. However, a simpler model can be derived while sacrificing very little accuracy.

A single-pixel SiPM is composed of an array of photodiodes that work near the breakdown region. A series-quenching resistor attached to every diode takes it quickly out of breakdown when needed (Geiger mode). Each photodiode with its quenching resistor is called a microcell. As a result, when a photon is absorbed by a microcell, a fast-current pulse of a fixed amplitude is generated. Consequently, the sum of all microcell outputs gives a pulse signal proportional to the number of activated microcells and is thus proportional to the number of photons absorbed by the SiPM. This behavior is modeled in Figure 5.2, where $R_q$ is the quenching resistor with its parasitic capacitance $C_q$, and $C_d$ and $C_g$ are the capacitances associated to the microcell and the connections. The values of these capacitances are in the range of 30–50 fF, which, due to the high number of microcells, will give capacitances up to hundreds of picofarads.

The avalanche mechanism that takes place in the diode is very fast, which makes it possible to model its contribution as a charge Dirac delta whose amplitude is

$$Q = \Delta V(C_d + C_q) \tag{5.11}$$

where $\Delta V$ is the overvoltage due to photon absorption in the Geiger-mode diode, which equals the difference between breakdown voltage and applied bias voltage. Thus, current generated by a microcell that absorbs a photon in time $t_0$ is

$$i(t) = Q\delta(t - t_0) \tag{5.12}$$

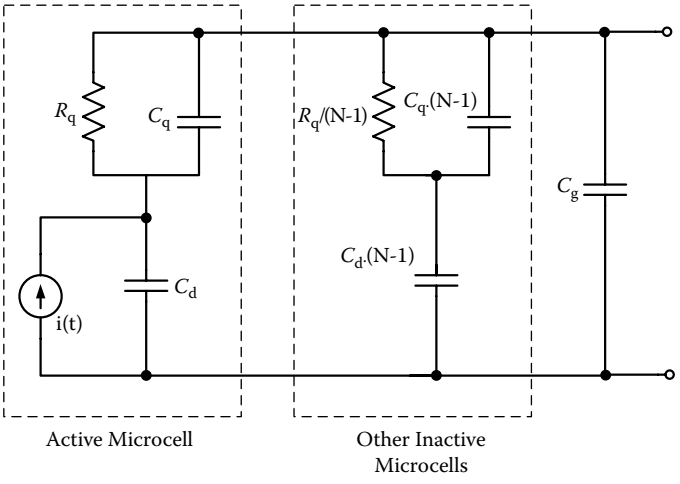



**FIGURE 5.2** Equivalent circuit of SiPM.

This model is valid for a single microcell activation that can be extended by superposition to describe SiPM behavior. Summing up all microcell contributions along a reduced period of time (a few nanoseconds), the pulse current is

$$i_{\text{total}}(t) = \sum_{n} Q\delta(t - t_n) \tag{5.13}$$

In the case where a current buffer with low input impedance is used as a front end, an equivalent circuit can be simplified, as shown in Figure 5.3. This approximation holds true if $R_p \ll (R_q/N)$, where $N$ is the number of microcells.

Repeating the same voltage analysis at the input of the front end that led to Equation (5.10), a new expression is obtained

$$V(t) = \frac{QR_p}{\tau_r - \tau_p}\left(\frac{\tau_q - \tau_p}{\tau_p} e^{-\frac{t}{\tau_p}} + \frac{\tau_r - \tau_q}{\tau_r} e^{-\frac{t}{\tau_r}}\right) \tag{5.14}$$

where $\tau_p = R_p(C_g + C_{eq})$, $\tau_r = R_q(C_d + C_q)$, $\tau_q = R_q C_q$, and $C_{eq} = \frac{(N-1)C_d C_q}{C_d + C_q}$. Also, $\tau_p$ and $\tau_r$ are the time constants of the preamplifier circuit and the subcell (also known as *recovering time*). Bearing in mind that the charge generation in a SiPM is almost instantaneous (depending solely on the avalanche process), it redistributes between $C_d$ and the series equivalent of capacitances $C_q$ and $C_{eq} + C_g$. As a consequence, at time zero

$$Q_p \cong Q\frac{C_q}{C_q + C_d} \;\rightarrow\; V(0) \cong \frac{Q_p}{C_{eq} + C_g} \tag{5.15}$$

However, this voltage value will be observable at amplifier output, depending on its bandwidth and slew-rate characteristics.

### 5.2.1.3 Direct Detector Equivalent Circuit

The morphology of direct detectors is simple enough to obtain a very compact equivalent model. However second-order effects, such as charge trapping and transit-time spread, strongly affect their behavior and cannot be well modeled. This fact makes it difficult to fully rely on electronic simulation data when dealing with these types of detectors.

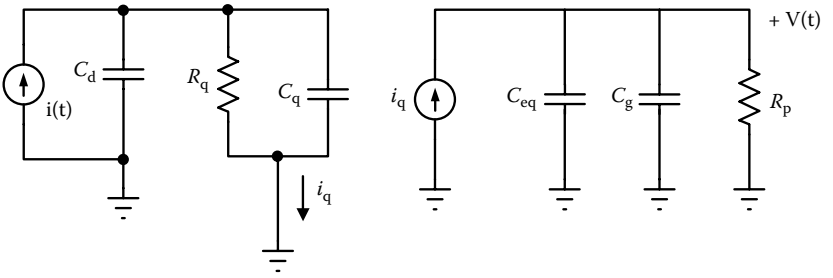



**FIGURE 5.3** Simplified equivalent circuit of SiPM connected to a current buffer.

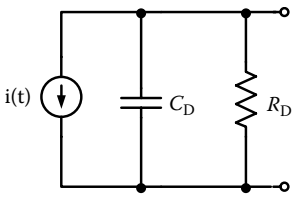



**FIGURE 5.4** Equivalent circuit of a direct detector.

A simple equivalent model such as that shown in Figure 5.4 can be applied to any semiconductor-based direct detector [7]. Capacitance $C_D$ and resistance $R_D$ correspond to that of the semiconductor crystal measured between anode and cathode. As in PSMT, the same model can be applied to every detector's output if a position-sensitive direct detector is being modeled.

The current pulse produced by the detector, modeled by $i(t)$ source, is generated by charge induction in detector terminals due to movement of charges (electrons and holes) obtained in the absorption process (Shockley-Ramo theorem [8, 9]). In order to obtain information about current expression, the first step is to calculate the movement of charges along the lines of the electric field applied to the detector. Electric field $\vec{E}$ can be obtained from applied voltage, the detector's geometry, and the resistivity of the semiconductor compound. Assuming $\overrightarrow{v_d}$ as a linear function of $\vec{E}$ and μ (charge mobility inside the semiconductor),

$$\overrightarrow{v_d} = \mu\vec{E} \;\rightarrow\; \overrightarrow{v_d} = \frac{d\overrightarrow{r(t)}}{dt} \;\rightarrow\; \vec{r} = \int \mu\vec{E}dt \tag{5.16}$$

As shown by the Shockley-Ramo theorem, given a weighting field associated with the detector's terminal (see Figure 5.5), the induced current is

$$i(t) = -q\overrightarrow{E_w} \cdot \overrightarrow{v_d} \tag{5.17}$$

To obtain total induced charge, Equation (5.17) has to be integrated for the time needed by the charge to get to the terminal (transit time). Nevertheless, Equation (5.17) shows that the contribution of a single charge and the number of charges obtained from an absorption process can be quite high. To calculate the full amount of the charge generated in a detector's output terminal, two approximations can be assumed. First, all of the charges are generated in the same point of the semiconductor; second, it takes the same time for all of the charges to get to the output terminal. That way, Equation (5.17) only has to be multiplied by the number of charges and by the mean transit time.

As noted previously, second-order effects can invalidate some of the assumed approximations. For example, some charges may get trapped if a semiconductor's crystal fails for a certain amount of time $\tau_D$. If this time is much smaller than the front end's collecting time, no effect will be observed. Otherwise, the collected charges will be less than those generated, and the detector's results may be affected.

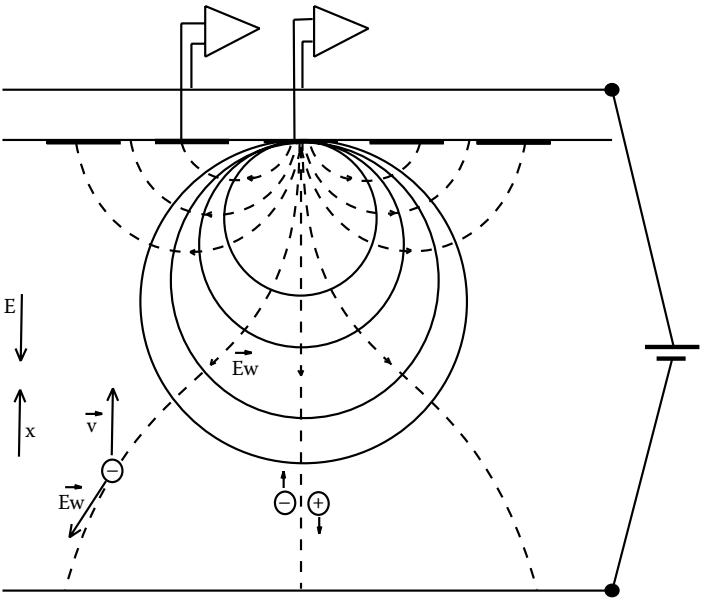



**FIGURE 5.5** Weighting field.

### 5.2.2 Effect of Signal Characteristics on Preamplifier Selection

The shape of the signal at the input of the preamplifying stage depends not only on detector characteristics, but also on the interaction between both. Changes in temporal characteristics of the pulses can be foreseen using equivalent circuits of the detectors and input impedance characteristics of the preamplifiers. In the specific case of direct detectors, where output signal magnitudes are so small and sensitive to external noise, charge amplifiers are used in order to collect every charge generated. These types of preamplifiers increase the charge measurement precision, but on the other hand, they change the shape of the pulse signal, modifying its temporal characteristics.

Depending on the specific application and detector type, two kinds of preamplifiers can be used: those that follow detector output as linear amplifiers and those that collect charge during a specific time window. The former ones will be named as *event-oriented preamplifiers*, while the latter will be named as *integrating preamplifiers*.

#### 5.2.2.1 Event-Oriented Preamplifiers

The main feature of these types of preamplifiers lies in their temporal response. A study by Knoll [2] covers this fact, although it is focused on current-to-voltage conversion with transresistance amplifiers. In this section, the discussion will be extended to current amplifiers, which show a more faithful response with lower distortion of input signal.

In the following discussion, the equivalent circuit of a PMT will be used, although the final conclusions apply to any kind of detector. In Equation (5.10), the $\lambda$ parameter corresponds to the inverse of scintillator decay time constant, and $a$ is the inverse

of the time constant of the whole circuit. Two different situations will be exposed, depending on whether θ is higher or lower than λ.

#### 5.2.2.1.1 High Circuit Time Constant

In the situation where $\theta \ll \lambda$, that is, when the circuit time constant is higher than the scintillator decay time constant, the original Equation (5.10) can be approximated as

$$V(t) \cong \frac{Q}{C}(e^{-\theta t} - e^{-\lambda t}) \tag{5.18}$$

Figure 5.6 shows a typical representation of a measured pulse under these conditions. Due to the high θ value in Equation (5.18), the second part of this expression dominates

$$\left(t \ll \frac{1}{\theta}\right) \quad \rightarrow \quad V(t) \cong \frac{Q}{C}(1 - e^{-\lambda t}) \tag{5.19}$$

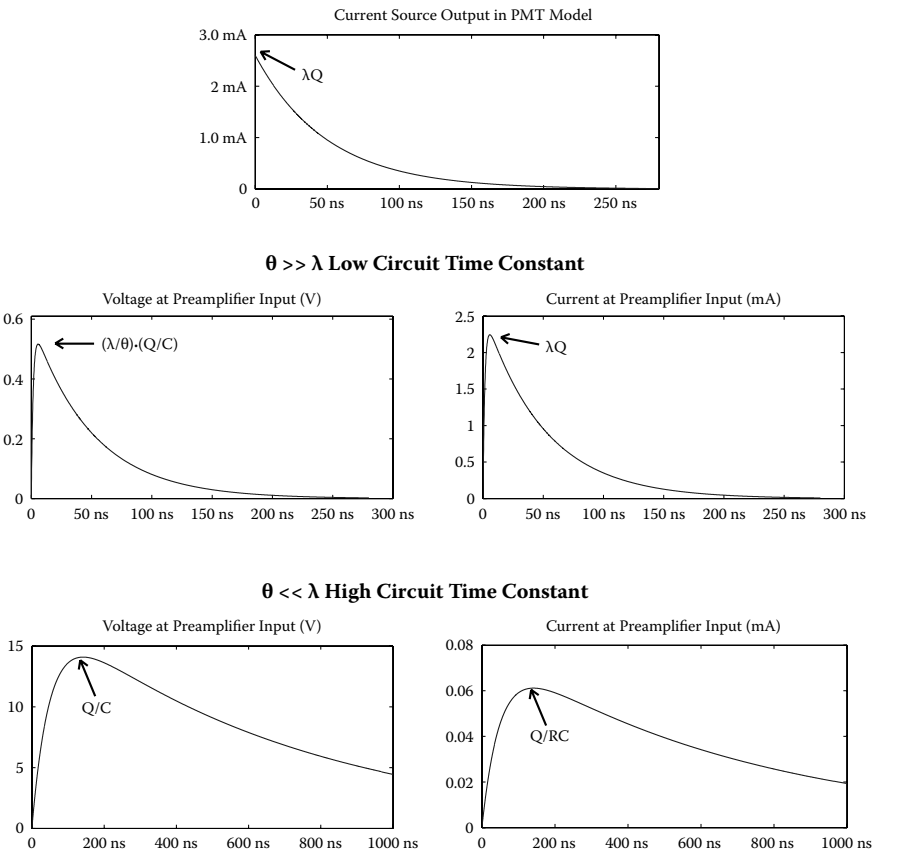



**FIGURE 5.6** Typical pulse signal in detector. Voltage and current at preamplifier input.

After a certain time, the total behavior is that of the first exponential term, as the second exponential has vanished

$$\left(t \gg \frac{1}{\lambda}\right) \quad \rightarrow \quad V(t) \cong \frac{Q}{C} e^{-\theta t} \tag{5.20}$$

In this situation, the rising part of the pulse is given by the time-decay constant of the scintillator. The faster the scintillator, the lower is the rising time of the output signal. On the other hand, the pulse tail is given by the circuit time constant (1/θ). The voltage pulse amplitude is given by $Q/C$, but it can only be achieved if $\theta \ll \lambda$. However, current pulse amplitude is given by $(Q/RC)$ under the same conditions.

Several conclusions can be stated from these results. The first is that the amplitude of the voltage signal is maximized at the expense of a longer pulse length, and the peak value does not directly depend on amplifier characteristics. The second important conclusion is that the current signal depends on the input impedance of the preamplifier. Increasing the pulse length increases the pileup probability, which can be very important at high input rates.

#### *5.2.2.1.2 Low Circuit Time Constant*

If the circuit time constant is lower than the scintillator decay time ($\theta \gg \lambda$), the situation is opposite to the previously discussed scenario. In this case, Equation (5.10) can be approximated as

$$V(t) \cong \frac{\lambda}{\theta} \frac{Q}{C} (e^{-\lambda t} - e^{-\theta t}) \tag{5.21}$$

Repeating the study in similar conditions, the first and the last part of the pulse are given by

$$\left(t \ll \frac{1}{\lambda}\right) \quad \rightarrow \quad V(t) \cong \frac{\lambda}{\theta} \frac{Q}{C} (1 - e^{-\theta t}) \tag{5.22}$$

$$\left(t \gg \frac{1}{\theta}\right) \quad \rightarrow \quad V(t) \cong \frac{\lambda}{\theta} \frac{Q}{C} e^{-\lambda t} \tag{5.23}$$

The corresponding current signals are

$$\left(t \ll \frac{1}{\lambda}\right) \quad \rightarrow \quad i(t) \cong \lambda Q (1 - e^{-\theta t}) \tag{5.24}$$

$$\left(t \gg \frac{1}{\theta}\right) \quad \rightarrow \quad i(t) \cong \lambda Q e^{-\lambda t} \tag{5.25}$$

Now the rise time depends on the time constant of the circuit, which in most cases is lower than the decay time of the scintillator. The pulse tail has the same response

as the detector's output pulse. The voltage peak value is quite low and depends on capacitance and preamplifier input impedance. Its value is ($\lambda Q/\theta C$), lower than that obtained with $\lambda \ll \theta$. However, if current is the magnitude chosen, the peak value is $\lambda Q$, which turns out to be the maximum value obtained from the detector.

Some conclusions can be stated from this study: Most detector applications show a low time-constant behavior. As a consequence, any signal processing at the output of the detector should be carried out in current mode to avoid signal-to-noise ratio losses. Any voltage conversion may need higher time constants in order to obtain higher amplitudes, thus degrading the temporal response. Unfortunately, a final current-to-voltage conversion is always needed before analog-to-digital conversion. At that point, there is a trade-off between voltage amplitude and pulse length. A long pulse leads to a higher amplitude, but on the other hand, it increases the pileup probability in high data-rate applications.

#### 5.2.2.2 Integrating Amplifiers

The use of direct detectors is always affected by two factors. First are the signal characteristics, which basically are charge pulses whose length is on the order of hundreds of nanoseconds or below. Second are the capacitive nature and its high equivalent impedance ($R_D$ has a very high value of 4). Under these conditions, front-end electronics must collect all of the charge generated (which is sometimes quite low) and turn it into voltage signals that can be further manipulated. Due to the capacitance of the detector, the preamplifier should show a very high equivalent capacitance, which would allow it to do the charge collection in an efficient way.

An integrating preamplifier based on a capacitive negative feedback is definitely the best choice for this situation. The design of the basic open-loop amplifier is of main importance to enhance the features of the final front end. Its gain should be high enough for the preamplifier to work with any detector capacitance. Also, the noise level should be low enough, since the amplifier will integrate this noise along with the charge signal.

The working principle of the integrating amplifier (Figure 5.7), also known as a charge amplifier, relies on small voltage variations at the input of the open-loop inverting amplifier. These small changes translate into large opposite voltage variations at the output, which allow charge collection in capacitor, $C_f$. The feedback mechanism guarantees a very low input voltage, which means that almost every charge generated by the detector will end up in $C_f$, independent of the $C_d$ value. Finally, $C_f$ shows a voltage proportional to the obtained charge, which turns out to be output voltage. Once the feedback capacitance is charged and the voltage is read, it has to be discharged to get ready for another input pulse. There are several ways to do this. The easiest way is to use $R_f$, which is constantly discharging $C_f$ at a much lower rate than the input pulse. It can affect pulse amplitude reading if the $R_fC_f$ time constant is not much longer than the pulse length. It will also take a lot of time to discharge $C_f$, which might be a serious issue with high rate inputs. There are several ways to overcome this problem [10], and these can be classified in two groups: The first group uses a switch to discharge $C_f$, which will need a synchronous mechanism to start and stop measurements. The second group uses continuous discharging techniques, sometimes based on low-frequency feedback loops to stabilize baseline voltage [11].

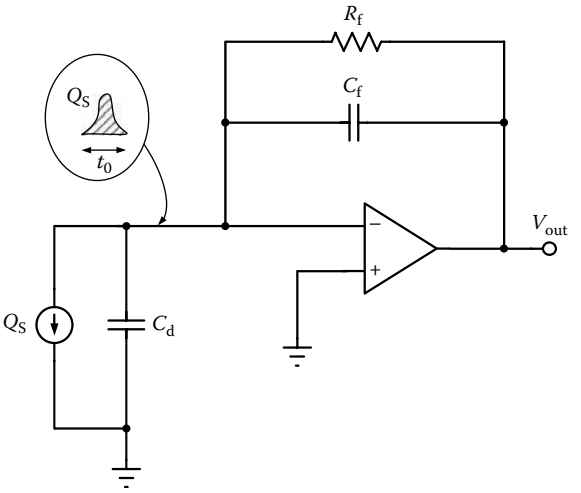



**FIGURE 5.7** Charge amplifier.

In order to obtain a detailed expression of output voltage, a Laplace transform can be applied to an ideal amplifier response. The integrating process between 0 and $t_0$ can be written as

$$Q_S(s) = Q_S\left(\frac{1}{s} - \frac{e^{st_0}}{s}\right) \tag{5.26}$$

The transfer function that establishes the relationship between the integrated charge and the output voltage can be modeled as an integrator with a first-order response

$$H(s) = -\frac{1}{C_f} \cdot \frac{1}{s + \frac{1}{\tau_f}} \tag{5.27}$$

where $\tau_f$ is the time constant of the feedback element ($\tau_f = C_f R_f$) (supposing a continuous discharge architecture). The output voltage expression in the Laplace domain is

$$V_{out}(s) = Q_S(s) \cdot H(s) = -Q_S\left(\frac{1}{s} - \frac{e^{-st_0}}{s}\right) \cdot \frac{1}{C_f} \cdot \frac{1}{s + \frac{1}{\tau_f}} = -\frac{Q_s}{C_f}\left(\frac{1}{s} \cdot \frac{1}{s + \frac{1}{\tau_f}} - \frac{e^{-st_0}}{s} \cdot \frac{1}{s + \frac{1}{\tau_f}}\right) \tag{5.28}$$

Applying an inverse Laplace transform to Equation (5.28), the output voltage in the time domain can be obtained for two different temporal segments:

$$V_{out}(t) = -\frac{Q_S}{C_f} \cdot \frac{1 - e^{-t/\tau_f}}{\left(t_0/\tau_f\right)} \qquad [0 \le t \le t_0] \tag{5.29}$$

$$V_{out}(t) = -\frac{Q_S}{C_f} \cdot \frac{e^{t_0/\tau_f} - 1}{\left(t_0/\tau_f\right)} \cdot e^{-t/\tau_f} \;\rightarrow\; (t_0 \ll \tau_f) \;\rightarrow\; V_{out}(t) = -\frac{Q}{C_f} e^{-t/\tau_f} \quad [t \ge t_0] \tag{5.30}$$

The results are quite similar to those obtained in event-oriented preamplifiers with high time constants. Although in this case, the maximum peak value depends on $C_f$ and not on total parasitic capacitance. This confirms an idea introduced later. Charge amplifiers are best used with detectors that generate low charge levels because the output voltage does not depend on parasitic capacitance.

### 5.2.3 Noise Analysis

The preamplifiers are important elements in the noise chain. As the first amplifying elements, they should have the best noise figure of the whole system in order to obtain a good signal-to-noise ratio. Noise will affect in a different way, depending on measured magnitudes and measurement methods. There are three main types of noise that have to be considered throughout the design process:

*Shot noise*: This noise has its origin in statistical charge fluctuations on a certain current signal. Current magnitudes are defined by a mean value of charges that flow at a certain time through a chosen measurement point plus a certain deviation that is given by Equation (5.31)

$$\overline{i^2} = \sqrt{2qI\Delta f} \tag{5.31}$$

Noise spectral density is constant ($\overline{i^2}/\Delta f$) and depends proportionally on current mean value I. This value can be calculated as the sum of a detector's dark current ID (mean current when no events are detected) plus preamplifier leakage current IL. Instantaneous amplitude of noise signal has a Gaussian shape distribution, so standard deviation is

$$\sigma = \sqrt{2qI\Delta f} \tag{5.32}$$

which means that, most of the time, noise amplitude is below $\pm 3\sigma$. This kind of noise is a major issue with bipolar devices.

*Thermal noise*: This kind of noise is associated with thermal movement of charges inside resistive elements, so equivalent current and noise quadratic values can be obtained. Noise spectral density is proportional to $kT$, where $k$ is the Boltzmann constant and $T$ is temperature:

$$\overline{v^2} = 4kTR\Delta f \tag{5.33}$$

$$\overline{i^2} = 4kT\frac{1}{R}\Delta f \tag{5.34}$$

As in shot noise, spectral density is constant and distribution is Gaussian shaped. This kind of noise can be quite dangerous when dealing with high resistive values or low resistive values associated with even lower or higher

input impedances. This combination makes the preamplifier system very sensitive to this type of noise.

*Flicker noise*: This noise is generated by defects and inhomogeneities in semiconductor crystals. Mosfet devices are very sensitive to this type of noise. Its most important characteristic is related to its frequency behavior, since its spectral density is proportional to 1/*f*. It is mainly associated with current magnitudes:

$$\overline{i^2} = K_1 \frac{I^a}{f} \Delta f \tag{5.35}$$

where $a$ is a constant in the range 0.5–2 and $K_1$ depends on device technology. In the CMOS process, pMOS devices usually have $K_1$ factors that are much lower than nMOS devices. Also, the $K_1$ constant is proportional to $1/WL$, which makes larger devices less noisy than smaller ones. These two reasons make pMOS devices a good initial choice for input stages, though other specifications such as gain or area must be fulfilled. Another important factor to bear in mind is that this kind of noise will affect low frequencies more intensively. The frequency where flicker noise reaches the thermal noise level is called the *corner frequency* and establishes the limit of flicker noise influence.

#### 5.2.3.1 Equivalent-Noise Charge Calculation

Noise effect on our system must be calculated in the design process to identify which parameters affect noise performance. Every detector together with its preamplifier configuration will have a certain noise response that will depend not only on total equivalent noise, but also on measurement methods. For instance, event oriented preamplifiers behave as classical linear amplifiers, so electronic noise contribution can be calculated as explained by Gray et al. [12]. Noise value in this cases corresponds to an uncertainty in a measured pulse amplitude. Integrating preamplifiers measure the total generated charge, so noise will also need to be integrated along the measurement time window.

Although conventional noise analysis can be applied, sometimes a more intuitive study is required to compare two different front ends. A general scheme with a simplified detector and front end is shown in Figure 5.8. The final goal is to obtain the different noise contributions as equivalent charge generated by the detector. This way, noise can be easily compared to signal magnitudes generated by the detector and also among different front-end configurations.

Shot noise from a detector appears in Figure 5.8 as source $i_{nd}$. Capacitance $C_d$ includes detector capacitance as well as any other parasitic capacitances present in the preamplifier input. Equivalent resistor $R_P$ represents detector impedance (in the case that a direct detector is used), possible biasing elements, and preamplifier input impedance. This resistance acts as a thermal noise source filtered by $C_d$. It is important to

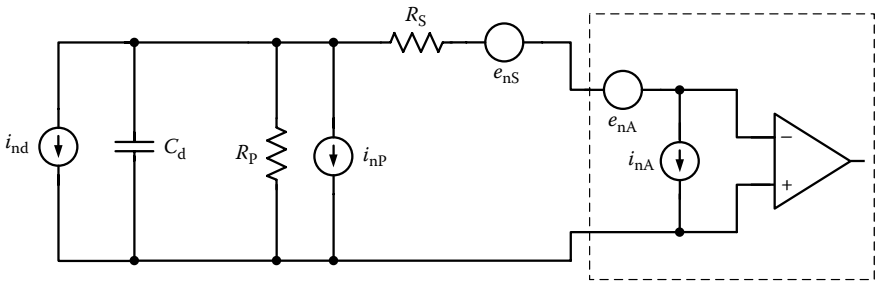



**FIGURE 5.8** Simplified model for equivalent noise charge calculations.

bear in mind that the quadratic mean of voltage in this case is independent of resistance value, since bandwidth is fixed by its own value:

$$\int_0^\infty e_P^2(\omega)d\omega = \int_0^\infty \frac{4kTR_P}{1+(\omega R_P C_d)^2}d\omega = \frac{kT}{C_d} \tag{5.36}$$

All series impedance present in the detector and preamplifier connection is modeled by $R_S$. Its thermal noise contribution is:

$$e_S^2(\omega) = 4kTR_S \tag{5.37}$$

Finally, $e_{nA}$ and $i_{nA}$ represent an equivalent-noise model of the preamplifier, which can be obtained analytically or by simulation. This source will include thermal, shot, and flicker noise contributions.

Equivalent-noise charge (ENC) is defined as the charge needed to obtain the same output signal in a noiseless model as that obtained from a noisy circuit. Following Radeka's work [13], ENC factor can be obtained as the sum of the variances of the outputs generated by each noise source under a normalized system response. The most important noise contributions are the following:

*Parallel white noise*: Corresponds to all the shot and thermal current noise sources. The expression of its contribution is

$$\mathrm{ENC}_{PW}^2 = \frac{1}{2}W_P\int_{-\infty}^{\infty} h^2(t)dt \tag{5.38}$$

where $W_P$ is the noise spectral density of each source, and $h(t)$ is the impulse response of the system. For example, for a detector shot noise source, $W_0 = 2qI$, the integral term can be approximated by a proportional estimation based on impulse response length $\tau$, where the $a_{Fl}$ factor depends on the exact shape of that response

$$I_P = \int_{-\infty}^{\infty} h^2(t)dt \simeq a_{F1}\cdot\tau \tag{5.39}$$

*Series white noise*: Is associated with the voltage noise source of the preamplifier and also with any series impedance between detector and preamplifier. Its expression is

$$\mathrm{ENC}_{\mathrm{SW}}^2 = \frac{1}{2} W_{\mathrm{S}} \int_{-\infty}^{\infty} [h'(t)]^2 dt \tag{5.40}$$

In this situation, the voltage mean square value has been transformed in a current mean square value. In order to do it, instead of using a Dirac delta train to model noise current, $\mathrm{d}'(t)C_{\mathrm{d}}$* functions will be used, so that finally $W_S = e_S^2 \cdot C_d^2$. The integral term can also be approximated as

$$I_{\mathrm{S}} = \int_{-\infty}^{\infty} [h'(t)]^2 dt \simeq \frac{a_{F2}}{\tau} \tag{5.41}$$

where the $a_{\mathrm{F2}}$ factor also depends on the exact shape of that response.

*Series flicker noise*: Convolution of 1/*f* noise source with impulse response generates a noise contribution that is independent of impulse response length [14]. This source contribution is given by Bertuccio, Pullia, and De Geronimo [15]:

$$\mathrm{ENC}_{\mathrm{SF}}^2 = a_{F3} \pi A_{\mathrm{f}} C_{\mathrm{d}}^2 \tag{5.42}$$

Finally, an ENC expression can be obtained summing up all noise contributions as

$$\mathrm{ENC}^2 = \tau \cdot a_{F2} \left(\frac{2kT}{R_{\mathrm{P}}} + 2qI_{\mathrm{L}} + i_{nA}\right) + \frac{a_{F1} C_{\mathrm{d}}^2}{\tau} \left(4kTR_{\mathrm{S}} + e_{nA}^2\right) + a_{F3} \pi A_{\mathrm{f}} C_{\mathrm{d}}^2 \tag{5.43}$$

## 5.3 INTRODUCTION TO CHARGE-DIVISION CIRCUITS

Position-sensitive particle detectors are indispensable for nuclear medical imaging and experimental particle physics. Charge division by means of resistive electrodes, resistive networks, and capacitive networks has been widely used for this purpose. It divides the ionization current or the photocurrent from a detector by having two paths to earth, with the resistance or capacitance of each path depending on the position of the electrical discharge.

The first detector using this principle was built by Wallmark [16] in 1957. He described a photocell able to detect the position of a small light spot at this cell. In 1978, Petersson and Lindholm [17] improved this photocell to measure not only the position of light spots on the *x*- or *y*-axis, but to allow also for a positioning of light spots at arbitrary locations on the detector surface. They also referred to Emmons [18], who pointed out that this lateral photovoltage had been observed by Schottky

* Convolution of δ′(t) with impulse response h(t) is h′(t).

in 1930 [19]. However, Schottky studied the origin of photoelectrons in photocells and did not recognize the practical significance of this effect for position detection. Alberi and Radeka [20] also attribute the first observation of charge division in resistive electrodes of photodetectors to Schottky.

The first time that charge division was used for position estimation in nuclear medicine was in 1958, when Anger developed the first scintillation camera [21]. Five years later, he presented the first positron camera [22]. Anger used both resistive and capacitive charge-division circuits for the position determination [23]. The use of charge division for particle detection with semiconductor devices was first proposed by Lauterjung et al. [24] in 1963. Subsequently, the method was also used with spark chambers by Charpak, Favier, and Massonnet [25] and with Geiger counters by McDicken [26] in 1963 and 1967, respectively. Since these initial applications, charge-division configurations have been used in a huge number of experiments, particle detectors, and medical-imaging devices. (See for instance [20, 27–29] and references therein.) Charge-division networks even found their way into artificial vision systems [30, 31] and were used as analog computers in solving complex boundary value problems [32, 33] and for analog simulation of physical media (see Vittoz [34] and references therein).

The signal propagation along the lossy transmission line of Figure 5.9 is well described by a coupled set of two first-order partial differential equations for the voltage $u(x,t)$ and the current $j(x,t)$, both functions of the position $x$ along the line and the time $t$.

$$\frac{\partial}{\partial x} j(x,t) = -\left[ c(x)\frac{\partial}{\partial t} + g(x) \right] u(x,t) = 0 \tag{5.44}$$

$$\frac{\partial}{\partial x} u(x,t) = -\left[ l(x)\frac{\partial}{\partial t} + r(x) \right] j(x,t) = 0 \tag{5.45}$$

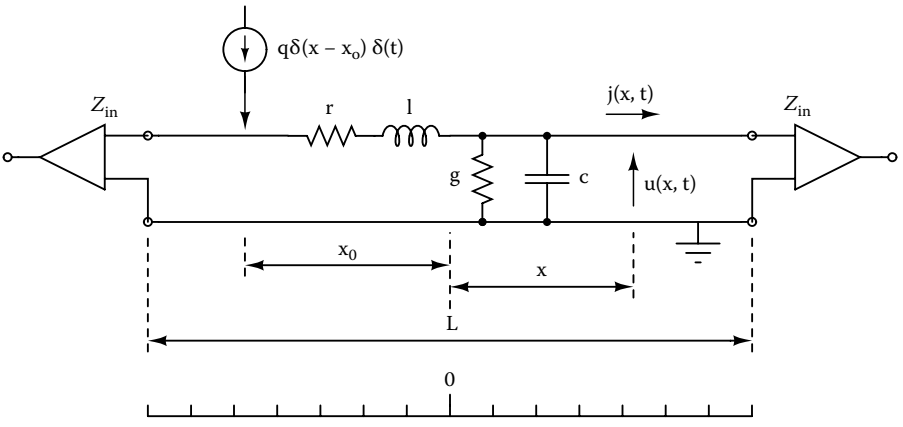



**FIGURE 5.9** Schematic representation of the components of a general transmission line.

In Equations (5.44) and (5.45), $r(x)$ and $l(x)$ are, respectively, the distributed resistance and inductance per unit length; $c(x)$ is the distributed capacitance between the conductors; and $g(x)$ is the conductance of the isolating material separating both conductors.

Differentiating Equations (5.44) and (5.45) with respect to $x$ and $t$ and by doing some manipulation, one obtains a set of two uncoupled second-order partial differential equations for the voltage $u(x,t)$ and the current $j(x,t)$. These equations are known as the telegrapher's equation.

$$\frac{\partial^2}{\partial x^2} j(x,t) = lc \frac{\partial^2}{\partial t^2} j(x,t) + (rc + gl) \frac{\partial}{\partial t} j(x,t) + grj(x,t) \tag{5.46}$$

$$\frac{\partial^2}{\partial x^2} u(x,t) = lc \frac{\partial^2}{\partial t^2} u(x,t) + (rc + gl) \frac{\partial}{\partial t} u(x,t) + gru(x,t) \tag{5.47}$$

The first term in Equations (5.46) and (5.47) represents electromagnetic wave propagation, the second term represents diffusion, and the third term represents attenuation. General solutions for Equation (5.47) in the case that the conductance $g(x)$ vanishes and a charge $q$ is injected at $x_0$, i.e., $u(x,0) = \frac{q}{c}\delta(x - x_0)$ can be found in Radeka and Rehak [35] and Masoliver, Porrà, and Weiss [36], among others. If the inductance $l(x)$ of the resistive electrode can also be neglected, Equations (5.46) and (5.47) become pure diffusive equations. Solutions for this case, have been given by Alberi and Radeka [20] and Borkowski and Kopp [28].

For practical particle detectors, one will not measure the signal distribution along the electrode, since this would lead to a huge amount of necessary analog-to-digital conversion (ADC) and therefore an expensive and complex data-acquisition system. Instead, one wants to encode a minimal number of signals with all relevant information, e.g., energy, position, transition time, and others. As will be seen, the currents $j(-L/2,t)$ and $j(-L/2,t)$ flowing out of the electrode at both ends can be used for this purpose. In many cases, photodetectors or particle detectors show a nontrivial transient response. For this reason, the asymptotic charges flowing out from each end are of special interest. These charges are obtained by integration of the currents at $x = -L/2$ and $x = -L/2$

$$Q_{\mathrm{l}} = \int_{t_0}^{\infty} j(-L/2,t)\,dt \tag{5.48}$$

$$Q_{\mathrm{r}} = \int_{t_0}^{\infty} j(L/2,t)\,dt \tag{5.49}$$

where $t_0$ is the starting moment of the injection of the charge with density $q(x)$. For the sake of simplicity and because the transient response of particle detectors and photodetectors are not objects of the present discussion, we assume that $q(t) = q\delta(t)$. For a maximum dynamic range of the position signal, one preferably adjusts the input impedance $Z_{\mathrm{in}}$ of the charge amplifiers at the ends of the line to be very low. This leads to the boundary conditions $u(-L/2,t) = u(L/2,t) = 0$. If the line was discharged

before the injection of the charge, then we obtain for the voltage distribution at this moment $u_0(x) = u(x,t = t_0) = q(x)/c(x)$. Furthermore, in all practical applications for medical imaging, the terms $g(x)u(x,t)$ and $l(x)\partial_t j(x,t)$ can be neglected. A vanishing conductance is preferred in order to achieve lossless charge collection. The inductance can be neglected, since it is very small for a charge divider for medical imaging. We then obtain, by integrating Equations (5.44) and (5.45) with respect to $x$ and $t$ and some manipulation [35], the following equation:

$$Q_{\mathrm{l}} = -\frac{\int_{-L/2}^{L/2} r(z)\left(\int_{-L/2}^{z} c(z')u_0(z')\right)dz'}{\int_{-L/2}^{L/2} r(x')dx'} \tag{5.50}$$

$$Q_{\mathrm{r}} = -\frac{\int_{-L/2}^{L/2} r(z)\left(\int_{z}^{L/2} c(z')u_0(z')\right)dz'}{\int_{-L/2}^{L/2} r(x')dx'} \tag{5.51}$$

For the special case that $r(x) = R/L = \text{const.}$, and since $u_0(x) = q(x)/c(x)$, we obtain

$$Q_{\mathrm{l}} = \frac{Q}{2} + \frac{Q}{L}\int_{-L/2}^{L/2} xq(x)dx \quad \text{and} \quad Q_{\mathrm{r}} = \frac{Q}{2} - \frac{Q}{L}\int_{-L/2}^{L/2} xq(x)dx \tag{5.52}$$

where $Q = \int_{-L/2}^{L/2} q(x)dx$ is the total charge that has been deposited on the wire. From Equation (5.52) it is easy to derive that

$$\frac{Q_{\mathrm{r}} - Q_{\mathrm{l}}}{Q_{\mathrm{r}} + Q_{\mathrm{l}}} = \frac{2}{L}\int_{-L/2}^{L/2} xq(x)dx \tag{5.53}$$

i.e., the charge divider computes the first moment (centroid) of the charge distribution $q(x)$ and simultaneously the total deposited charge $Q$. The centroid is also known as the center of gravity.

Gaseous detectors with resistive wire electrodes are rarely used for γ-ray imaging due to their low detection efficiency. Typical γ-photon energies in medical applications are 140 keV and 511 keV. Detectors build with scintillation crystals and segmented photodetectors are generally used for these energies. A general discussion of a discrete capacitive and resistive charge-division circuit was given by Pullia et al. [37]. The discrete resistive charge divider is shown in Figure 5.10. $C_{\mathrm{p}}$ is the (parasitic) capacitance of one segment of the position-sensitive photodetector. The whole charge divider can be considered as a chain of $n_{\mathrm{l}} + n_{\mathrm{r}}$ elementary cells connected to the outputs of a photodetector with a one-dimensional array of $n$ pixels. Each elementary cell (dotted box in Figure 5.10) consists of a resistance $R$ and two capacitors with half the capacitance of the photodetector segment. The

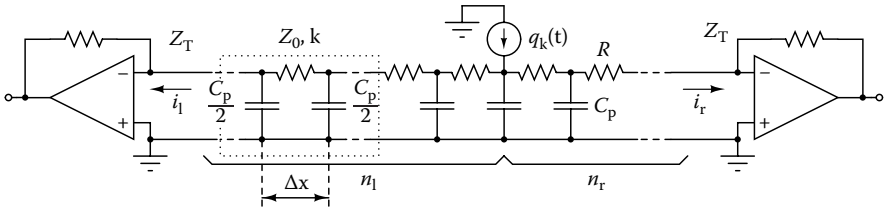



**FIGURE 5.10** Schematic representation of the components of a general transmission line.

current $j_r(t)$ that will be observed at the input of the right amplifier ($Z_T \approx 0$) if the charge $q_k(t) = q_k\delta(t)$ is injected at position $k$ can be derived to be

$$j_r(k,s) = j_k \frac{\sinh\left[k\cosh^{-1}\left(1+\frac{1}{2}sRC_p\right)\right]}{\sinh\left[(n_l+n_r)\cosh^{-1}\left(1+\frac{1}{2}sRC_p\right)\right]} \tag{5.54}$$

in the frequency domain, with $s$ being the Laplace variable. Equation (5.54) can be shown to only have simple poles. Therefore, the Heaviside expansion theorem (see, e.g., Davies [38]) can be used to perform the inverse Laplace transform for obtaining the transient response of the discrete transmission line

$$j_r(k,t) = \frac{2q_k}{(n+1)\tau}\sum_{h=1}^{n-1}(-1)^{h+1}\sin\left[\frac{kh\pi}{n+1}\right]\sin\left[\frac{h\pi}{n+1}\right]e^{\frac{-4t}{\tau}\sin^2\left[\frac{h\pi}{2n+2}\right]} \tag{5.55}$$

where $n_r$ and $n_l$ are clear from Figure 5.10, $n + 1$ is equal to $n_l + n_r$, and $\tau = RC_p$ is the time constant of a single cell of the charge divider (dotted frame in Figure 5.10). The current to the left is obtained by substitution of $k$ by $n + 1 - k$. Pullia et al. [37] found that the charge transfer of the corresponding fraction of $q_k$ is not complete before approximately $0.5n^2\tau$. The charge collection at the right amplifier for a divider with 10 cells is shown in Figure 5.11 to the left. The charge loss due to incomplete integration of the current in Equation (5.55) is shown in the same figure to the right. In the case that the data-acquisition system is built with charge-integrating ADCs, the network's processing time of $\Delta_{proc} \approx 0.5n^2\tau$ does not introduce significant errors as long as the integration time is larger. However, free-running sampling data acquisition systems will introduce important positioning errors when the position is derived from the maximum of the charge pulse (see Herrero et al. [39]).

## 5.4 IMPLEMENTATION OF CHARGE-DIVISION CIRCUITS

A one-dimensional implementation for a discrete charge divider is therefore given by a chain of resistors, as shown in Figure 5.10. Integration of the currents $j_r(k, t)$ and $j_l(k, t)$ with respect to time significantly simplifies Equation (5.55) and gives the following result for the relative position of the single charge $q_k$ at position $k$:

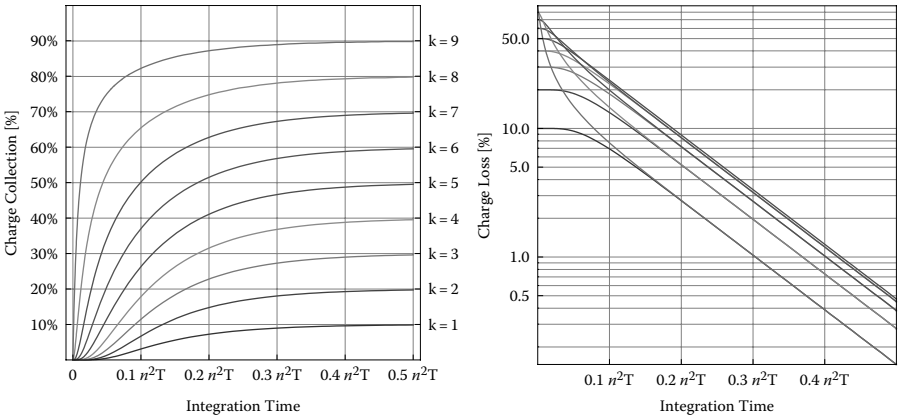



**FIGURE 5.11** Left: Charge transfer time dependency to the right end of the discrete transmission line of Figure 5.9. Right: Errors in the case of incomplete charge collection, i.e., insufficient integration time.

$$Q_r(k) = \int_0^\infty j_r(k,t) = q_k \frac{k}{n+1}, \quad Q_l(k) = \int_0^\infty j_l(k,t) = q_k \frac{n+1-k}{n+1} \tag{5.56}$$

and

$$\frac{Q_r(k) - Q_l(k)}{Q_r(k) + Q_l(k)} = \frac{2k - n - 1}{n+1} \tag{5.57}$$

Thus, if we assume a charge-integrating data-acquisition system with an integration time larger than $0.5n^2\tau$, we can neglect the effects from the capacitances $C_p$, and the charge divider behaves like a simple resistor chain with static currents $J_k$, $J_r(k)$, and $J_l(k)$. If several charges $q_k$ (or currents) at distinct positions $k$ are injected into the divider, the currents will be superposed and we obtain for the position

$$\frac{Q_r - Q_l}{Q_r + Q_l} = \frac{2}{n+1} \frac{\sum_{k=1}^{n} k q_k}{\sum_{k=1}^{n} q_k} - 1 = \frac{2}{n+1} \frac{\sum_{k=(1-n)/2}^{(n-1)/2} k J_k}{\sum_{k=(1-n)/2}^{(n-1)/2} J_k} = \frac{J_r - J_l}{J_r + J_l} \tag{5.58}$$

where we shifted the index $k$ by $-(1 + n)/2$ in the next-to-last step. The quotient of sums is just the centroid of the currents $J_k$ that are injected at the inputs of the divider. If we take the pitch $\Delta x$ of the pixels into account, we finally get

$$\langle x \rangle = \frac{\sum_k x_k J_k}{\sum_k J_k} = (n+1) \frac{\Delta x}{2} \frac{J_r - J_l}{J_r + J_l} \tag{5.59}$$

The circuit in Figure 5.10 can be used with multiwire anode position-sensitive photomultipliers. Photodetectors with a regular matrix of detector segments require adaptation of the charge-division principle to the two-dimensional case. Anger proposed to use four resistors for each pixel $(i, j)$ with appropriate values. The current fractions of the input current $J_{ij}$ will be fed into four independent buses $J_l$, $J_r$, $J_u$, and $J_d$, as in Figure 5.12, that are connected to a charge amplifier with low input impedance. The $x$ and $y$ centroids can then be computed by virtue of

$$\langle y\rangle = \frac{J_u - J_d}{J_u + J_d} = \frac{\sum_{i,j} \frac{R^d_{i,j} - R^u_{i,j}}{R^d_{i,j} + R^u_{i,j}} J_{i,j}}{\sum_{i,j} J_{i,j}} \quad \text{and} \quad \langle x\rangle = \frac{J_r - J_l}{J_r + J_l} = \frac{\sum_{i,j} \frac{R^r_{i,j} - R^l_{i,j}}{R^r_{i,j} + R^l_{i,j}} J_{i,j}}{\sum_{i,j} J_{i,j}} \tag{5.60}$$

where the resistor values are given by

$$R^l_{i,j} = \frac{a_{i,j}}{1 - x_i}, \quad R^r_{i,j} = \frac{a_{i,j}}{1 + x_i}, \quad R^u_{i,j} = \frac{a_{i,j}}{1 + y_j}, \quad \text{and} \quad R^d_{i,j} = \frac{a_{i,j}}{1 - y_j} \tag{5.61}$$

Note that there is no constraint for the $a_{ij}$. They actually can have any value without affecting the correct charge division.

Borkowski and Kopp [28] and Siegel et al. [29] proposed a multidimensional charge divider based on resistor chains, as shown in Figure 5.12 to the right. This network can be used with a position-sensitive photodetector with $n \times m$ outputs, and a circuit analysis of it gives that the $x$ and $y$ position can be computed by

$$\langle x\rangle = \Delta x \frac{n+1}{2} \frac{J_b + J_c - J_a - J_d}{J_a + J_b + J_c + J_d}, \quad \langle y\rangle = \Delta y \frac{m+1}{2} \frac{J_a + J_b - J_c - J_d}{J_a + J_b + J_c + J_d} \tag{5.62}$$

with $\Delta x$ and $\Delta y$ being the pitch of the detector segments in the $x$- and $y$-directions, respectively. For correct positioning, the horizontal resistors of the first and last columns, i.e., $R_{hl}$, have to fulfill the condition $R_{hl} = R_h + \frac{1}{2} l(l - m - 1)R_v$ with $1 \in [1, 2, \ldots, m]$ [29].

Hybrid solutions that combine Anger's approach and one-dimensional resistor chains are also possible. Siegel et al. [29] proposed a hybrid network that performs the primary charge division by means of a one-dimensional resistor chain and the secondary division by a one-dimensional Anger logic. This topology is shown in Figure 5.13. For this implementation, the positions can be computed using Equation (5.62), if the resistances at the end of the vertical chains are given by

$$R^u_j = \frac{R_h}{\frac{1}{2} - \frac{j}{m+1}} \quad \text{and} \quad R^d_j = \frac{R_h}{\frac{1}{2} + \frac{j}{m+1}} \tag{5.63}$$

Popov, Majewski, and Welch [41] and Olcott et al. [42] use the reverse order by first applying a one-dimensional Anger logic that feeds the current fractions through a buffering amplifier into a one-dimensional resistor chain, as shown in Figure 5.14.

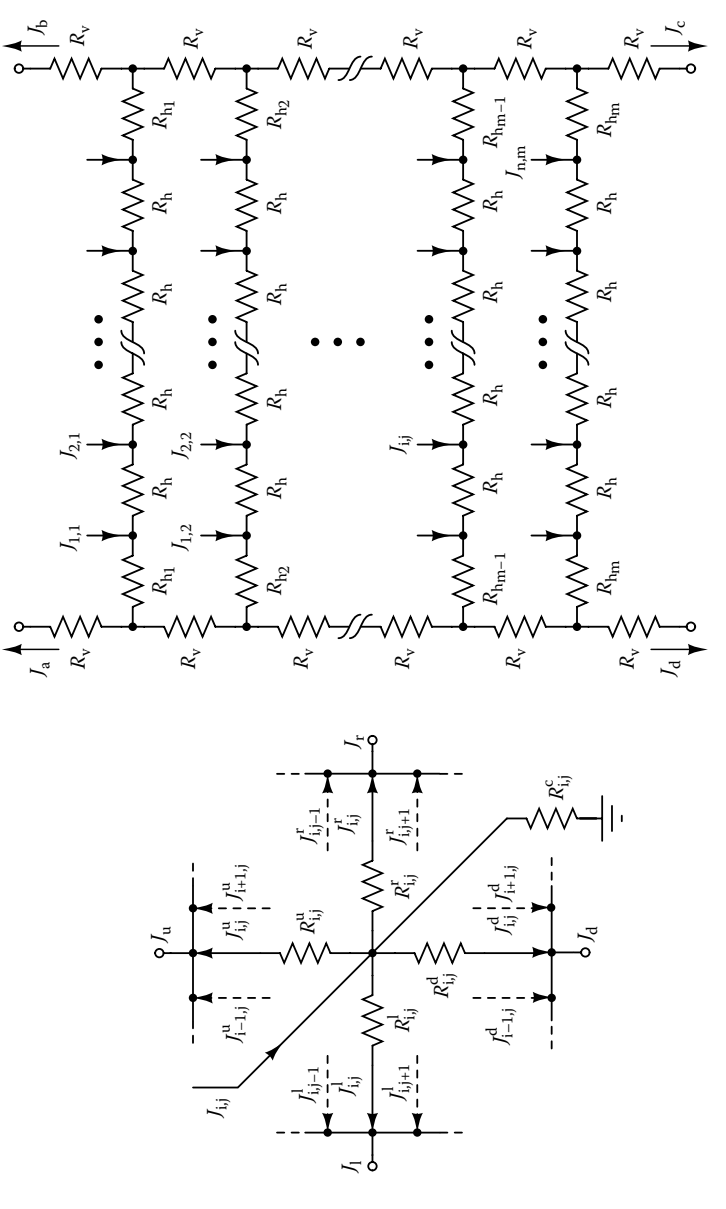



**FIGURE 5.12** Left: Positioning logic as proposed by Anger for the scintillation camera. The resistor $R^c_{i,j}$ is optional; its purpose is to correct the unequal signal strength from different pixels [40] in the case that the original Anger network is used. Right: Two-dimensional proportional resistor network for a position-sensitive photodetector with an $n \times m$ matrix of outputs.

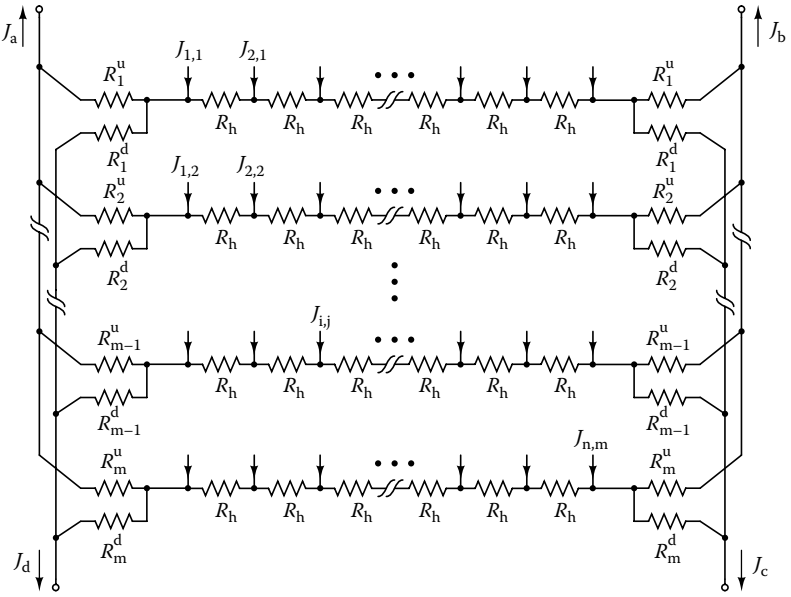



**FIGURE 5.13** Circuit for the two-dimensional hybrid charge dividing readout. It uses $m$ horizontal resistor chains for the $x$-centroid and Anger-type charge division scheme for the other spatial direction.

As in the case of the $a_{i,j}$ for the Anger logic, there is no constraint for the resistor values $R_{i,j}$ of the primary divider. Using this topology, the positions can be computed by means of Equation (5.60).

The difference in the positioning behavior of the first three solutions (Figures 5.12 and 5.13) was studied by Siegel et al. [29]. They found their performance to be very similar. As for the second hybrid solution, Olcott et al. [42] report that the dynamic ranges of the charge divider's output currents are approximately five times larger for the network that uses exclusively proportional resistor chains (Figure 5.12, right) when compared to the fourth solution (Figure 5.14), while its root mean square noise is only half as large as for the solution by Olcott et al.

An advantage is the fact that several distinct position-sensitive photodetectors can combined with a single charge-division circuit. (See for instance Seidel et al. [43] and Zhang et al. [44].) This allows a significant reduction in the number of signals to process and hence simplifies electronics complexity. The disadvantage of connecting several position-sensitive photodetectors to a single network is that the count-rate independence of each photodetector is lost. The event rate seen by the ensemble of photodetectors connected to a single charge-division circuit is the sum of event rates seen by each single detector, raising the probability of pulse pileup.

It was also found that all discrete charge dividers discussed so far can be extended to allow measuring the standard deviation of the current/charges from the photodetector segments [1, 45]. For the illustration of this property, consider the one-dimensional proportional resistor chain as shown in Figure 5.15 for a one-dimensional detector with $n$ detector outputs. If a single current $J_k$ is injected at an arbitrary node

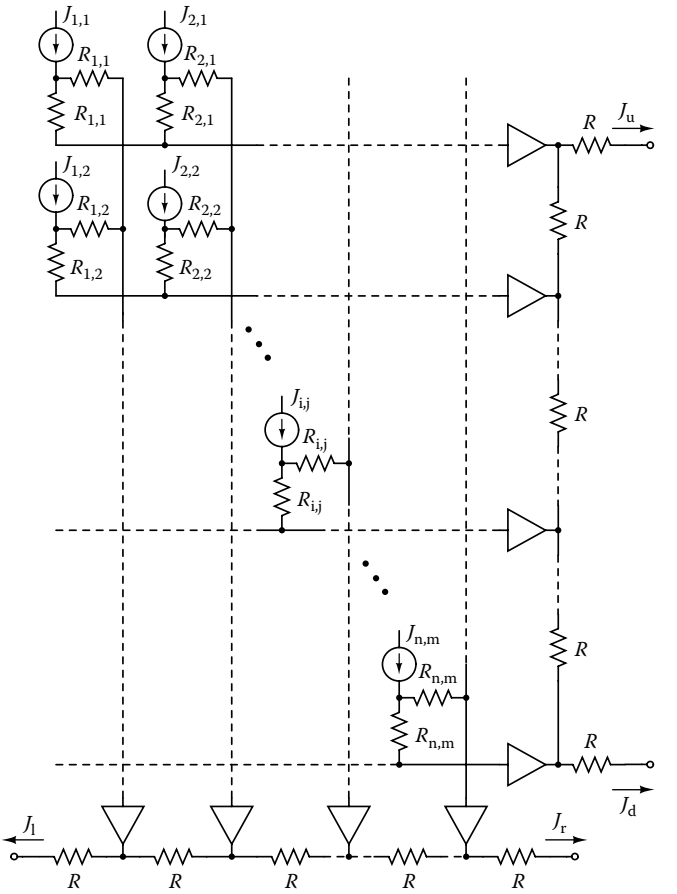



**FIGURE 5.14** Alternative hybrid implementation of a two-dimensional hybrid charge dividing readout.

of index $k \in \left[\frac{1-n}{2}; \frac{n-1}{2}\right]$, it will generate voltages $U_{k,o}$ at all nodes $o \in \left[\frac{1-n}{2}; \frac{n-1}{2}\right]$. The voltage $U_o$ observed at position $o$ is the superposition of all $U_{k,o}$, i.e., $U_o = \Sigma_k U_{k,o}$. One observes that the current $J_k$ at the point $x_k$ will sustain an impedance

$$Z_k = R_k^l \,||\, R_k^r = \frac{R_d}{n+1}\left(\frac{(n+1)^2}{4} - k^2\right) \tag{5.64}$$

where $R_k^l = R_d\left(\frac{n+1}{2} + k\right)$ and $R_k^r = R_d\left(\frac{n+1}{2} - k\right)$ are the total resistances to the left and to the right. The voltage $U_{k,k}$ seen at this same position is simply the product of current $J_k$ and the impedance $Z_k$:

$$U_{k,k} = \frac{J_k R_d}{n+1}\left(\frac{(n+1)^2}{4} - k^2\right) \tag{5.65}$$

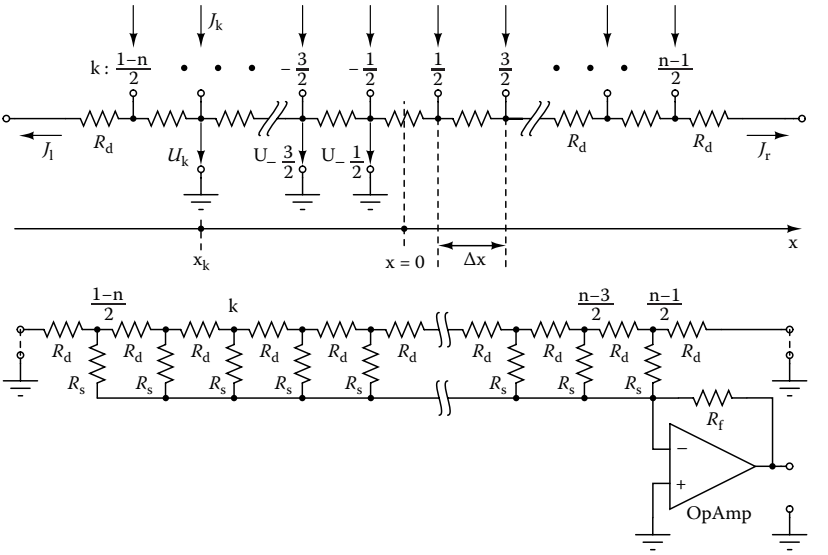



**FIGURE 5.15** Top: One-dimensional proportional resistor chain and naming conventions. Bottom: One-dimensional proportional resistor chain with summing amplifier for second-order moment computation.

The required quadratic weights are therefore inherently available within this charge-division circuit. However, an attached analog adder, as in Figure 5.15 (below), does not give the sum of all $U_{\mathrm{k,k}}$ generated by the currents at different $i$ because the resistor chain acts as a voltage divider and induces fractions of $U_{\mathrm{k,k}}$ at all $o$ with $o \neq k$. Since all resistors in the chain have the same value $R_{\mathrm{d}}$, the voltage fractions $U_{\mathrm{o\leq k,o}}$ and $U_{\mathrm{o>k,o}}$ to the left and right of the injection position $i$ are given by

$$U_{o\leq k,o} = U_{k,k}\frac{1+2o+n}{1+2k+n} \quad \text{and} \quad U_{o>k,o} = U_{k,k}\frac{1-2o+n}{1-2k+n} \tag{5.66}$$

As a consequence of the superposition principle, the adder sums up all of the voltage fractions $U_{\mathrm{k,o}}$ (including $U_{\mathrm{k,k}}$) of all $J_{\mathrm{k}}$. As long as $R_s \gg R_k^l \left\| R_k^r \forall k \in \left[\frac{1-n}{2}; \frac{n-1}{2}\right]\right.$, each summand $U_{\mathrm{k,o}}$ will be amplified by the weight $-R_{\mathrm{f}}/R_{\mathrm{s}}$, and gives rise to the output voltage

$$U_\Sigma \approx -\frac{R_f}{R_s}\sum_k^n \sum_o^n U_{k,o} \tag{5.67}$$

where the sum over $o$ is given by

$$\sum_o U_{k,o} = U_{k,k}\left(\sum_{o=\frac{1-n}{2}}^{k} \frac{1+n+2o}{1+n+2k} + \sum_{k+1}^{o=\frac{n-1}{2}} \frac{1+n-2o}{1+n-2k}\right) = U_{k,k}\frac{n+1}{2} \tag{5.68}$$

This result is important, since it states that the contribution of this partial sum to the complete sum of Equation (5.67) reduces to the same constant factor $(n+1)/2$ independently of the position $k$, where the current $J_k$ is injected. Using Equation (5.65) and evaluating the sum over $k$ gives

$$\frac{U_\Sigma}{J} \approx \frac{g_\Sigma R_f R_d}{2 g_J R_s} \left( \frac{(1+n)^2}{4} - \frac{\sum_k k^2 J_k}{\sum_k J_k} \right) \tag{5.69}$$

where $g_J$ is the gain for the electronic channels for the currents $J_r$ and $J_l$, and $g_\Sigma$ is the gain for the electronic channel for the sum $U_\Sigma$. By solving the equation for the second-order moment and using $x_k = k\Delta x$, one obtains

$$\left\langle x^2 \right\rangle = \frac{\sum_k x_k^2 J_k}{\sum_k J_k} \approx \Delta x^2 \left( \frac{(1+n)^2}{4} - \frac{2 g_J R_s}{g_\Sigma R_f R_d} \frac{U_\Sigma}{J_r + J_l} \right) \tag{5.70}$$

Finally, $\sigma^2 = \langle x \rangle^2 - \langle x \rangle^2$ can be computed from Equations (5.59), (5.69), and (5.70). This result is of special interest for positron emission tomography. The rather high energy of the annihilation photons requires thick scintillation crystals, even in the case that scintillators with high density and effective atomic numbers are used. Although it is possible to estimate the positions parallel to the photodetection plane with charge-division circuits, the normal component of the photoconversion position cannot be obtained without additional information. Since it has been observed by different groups, e.g., Rogers et al. [46] and Matthews et al. [47], that this third component and the width of the scintillation light distribution are strongly correlated, Equation (5.70) can be used to estimate the depth of interaction of the γ-ray. Note, however, that Equation (5.70) requires particle detectors with monolithic scintillation crystals, while the position measurement also works with an array of small scintillator pixels.

## 5.5 SYSTEMATIC AND STATISTICAL ERRORS

Important sources for systematic errors are discretization effects and, in the case of monolithic scintillators, border effects. The first are a consequence of the sampling of an arbitrary signal distribution that enforceably includes its discretization. Almost always, this destroys existing symmetries of the distribution. Border effects arise because it is impossible to realize detectors of infinite spatial extension, and this also leads to breaking of the symmetries of the signal distribution. On the other hand, one faces (a) signal fluctuations that are inherent to the detection process and (b) in the case that a scintillator and a photodetector are used, a quantum detection efficiency smaller than 100%.

Fluctuations in the signal distribution are caused by various underlying processes. In the case of a scintillation detector, the signal generation starts with the photoconversion of the γ-ray into one or more electrons within the scintillation crystal.

Depending on the underlying process, e.g., the Compton effect, photoelectric effect, or pair production, one or more primary electrons are generated. These primary electrons produce the scintillation light via the decay of excited scintillation centers. Together with these random processes, inhomogeneities of the material and nonproportional energy dependency lead to important fluctuations in the total light output of the scintillator. While this affects all detector segments by the same amount, the distribution of the finite number of these scintillation photons over the sensitive area of the detector leads to fluctuations that differ for all detector segments. When the scintillation light arrives at the photodetector, the photons will be converted independently and one by one to photoelectrons. This is a Poisson process and probably the most important cause for segment-dependent signal fluctuations for photomultipliers. Finally, the fluctuations will lead to uncertainties $d_P\langle x\rangle$ and $d_P\langle y\rangle$ in the centroid measurements $\langle x\rangle$ and $\langle y\rangle$, which can be estimated by error propagation. The subscript in $d_P\langle x\rangle$ indicates that this error contribution is due to the Poisson nature of light detection. Consider a set $\{q_k\}$, $k = [1, 2, \ldots, n]$ of $n$ charges that represent the signals of the $n$ detector segments. Straightforward computation gives for the $x$-centroid error

$$\delta_P\langle x\rangle = \sqrt{\sum_k \left[\frac{\partial}{\partial q_k}\left(\frac{\sum_k x_k q_k}{\sum_k q_k}\right)\delta q_k\right]^2} = \frac{1}{\sqrt{\sum_k q_k}}\sqrt{\frac{\sum_k (x_k - \langle x\rangle)^2 \delta q_k^2}{\sum_k q_k}} \tag{5.71}$$

where the centroid is given substituting the currents $\{J_k\}$ by the set of charges $\{q_k\}$ in Equation (5.59). Due to the Poisson nature, the fluctuation by $\delta q_k$ is proportional to $\sqrt{q_k}$, and we can write $\delta q_k = \alpha\sqrt{q_k}$, where $\alpha$ is the proportionality constant characterizing the photodetector. Inserting this in Equation (5.71), we obtain

$$\delta_P\langle x\rangle = \frac{\alpha}{\sqrt{\sum_k q_k}}\sqrt{\frac{\sum_k (x_k - \langle x\rangle)^2 q_k}{\sum_k q_k}} = \alpha\frac{\sigma}{\sqrt{Q}} \tag{5.72}$$

with $\sigma$ being the standard deviation of the input distribution and $Q$ being the integral input charge.

This result is not unexpected. Since the standard deviation gives an idea of the dispersion of the set of variables $\{q_k\}$, Equation (5.72) states that the error of the centroid decreases as the signal distribution narrows. Furthermore, the error also scales with the square root of the sum of all signals $\{q_k\}$, e.g., the total amount of charge produced by the incoming scintillation light photons. One therefore expects a better position estimate from the center-of-gravity algorithm for scintillators with high light yield and narrow light distributions.

For the discrete implementation of charge-division circuits, the set of $n$ signals $\{q_k\}$ has to be provided by a photodetector that samples the scintillation light distribution. Normally, this sampling is done by integrating the distribution piecewise

over many small and equidistant intervals, where these intervals correspond to the detector segments of size ρ.

$$q_k = \int_{\rho(n-\frac{1}{2})}^{\rho(n+\frac{1}{2})} \varphi(x)dx \tag{5.73}$$

This sampling can be considered as a convolution [48] of the real distribution $\varphi(x)$ with the boxcar function $\Pi(x)$ of width ρ leading to a new, effective distribution:

$$q_{\text{eff}}(x) = \int_{-\infty}^{\infty} \Pi\left(\frac{x-\zeta}{\rho}\right)\varphi(\zeta)d\zeta \tag{5.74}$$

The set of input charges values $\{q_k\}$ is then given by the value of $q_{\text{eff}}(x)$ at the positions of the pixels $x = x_k$. i.e., $q_k = q_{\text{eff}}(x_k)$. It is clear from Equation (5.74) that the resulting distribution $q_{\text{eff}}(x)$ has to be wider than the original distribution $\varphi(x)$, since one always has $\rho > 0$ for any real photodetector. As a consequence, the pixel size of the photodetector can also limit the spatial resolution in the case that monolithic scintillation crystals are used. In the case of segmented scintillators, the intrinsic spatial resolution is given by the size of the scintillator segment. As an example, we consider the projection of the normalized one-dimensional inverse-square law onto the abscissa $x$,

$$\varphi(x) = \frac{dJ}{\pi(x^2+d^2)} \tag{5.75}$$

to study the influence of the detector pixel size ρ on the width of the sampled distribution $q_{\text{eff}}(x)$. While the undisturbed distribution in Equation (5.75) has a full width at half maximum (FWHM) width equal to $2d$, the sampled distribution $q_{\text{eff}}(x)$ will have the width $2\sqrt{\rho^2/4+d^2}$. For small $d$ and small ρ, the standard deviation of $q_{\text{eff}}(x)$ can be estimated to be

$$\sigma_S = \delta_S\langle x\rangle \approx \frac{\rho^2}{12} + \frac{dL}{12\pi}\left(12 - \frac{8\rho^2}{L^2}\right) \tag{5.76}$$

and represents the centroid error due to finite size effects of the detector and its pixels. Equation (5.76) depends on the size $L$ of the detector, because this parameter defines the lower and upper limit for the computation of an arbitrary moment $\mu_k$ of the distribution $\varphi(x)$

$$\mu_k(x) = \int_{-L/2}^{L/2} x^k\varphi(x)dx \tag{5.77}$$

(Note that the relation FWHM = $2.35\sigma$ cannot be applied to $2\sqrt{\rho^2/4+d^2}$, since it only holds for Gaussian distributions.) Equation (5.77) only leads to error-free moments for the case that the integration interval $[-L/2, L/2]$ covers the support of

the distribution, which is not the case for the inverse-square law. Apart from the error observed at the center of the detector from Equation (5.76), one will find additional discretization errors for the case that the unimodal distribution does not take its maximum at the center of the interval [−*L*/2, *L*/2]. These errors have been studied by Landi [48], who found that the centroid can be written as

$$\langle x \rangle = x_0 + \frac{\rho}{\pi} \sum_{h=1}^{\infty} \frac{(-1)^h}{h} \sin\left[\frac{2\pi h x_0}{\rho}\right] \Phi\left[\frac{2\pi h}{\rho}\right] \tag{5.78}$$

where F(ω) is the Fourier transform of φ(*x*), and $x_0$ is the position of the maximum of the distribution and hence the position of the γ-ray photoconversion. From the Fourier series of the centroid discretization error, it can be seen that the amplitudes $(-1)^h F(2\pi h/\rho)/h$ of the spatial frequencies $\omega_h = hx_0/\rho$ scale with the sampling interval ρ. Evidently, the discretization error completely vanishes for all $x_0 = \rho n/2$, that is, for all those cases when $x_0$ is located exactly over the center of one pixel or exactly between two pixels. At these special points, the symmetry of the distribution will be correctly reproduced by the detector (disregarding the effect of the finite dimension *L*). If φ(*x*) converges to the Dirac d-function, the discretization errors reach their maximum values. This is intuitively expected because the exact position of the Dirac d-function over one pixel cannot be determined. We can compute the Fourier transform of the sample distribution given in Equation (5.75) and obtain for the amplitudes the values

$$\Phi(2\pi h/\rho) = \frac{e^{-\frac{2dh\pi}{\rho}} J}{\sqrt{2\pi}} \tag{5.79}$$

With this result, one can easily find an upper limit for the discretization error in Equation (5.78), since $e^{-2d\pi/\tau}/\sqrt{2\pi}$ emphasizes the series in Equation (5.78). Computing the alternating harmonic series, one obtains the following equation:

$$\delta_D \langle x \rangle \lesssim \frac{e^{-\frac{2d\pi}{\tau}} \tau \ln(2)}{\sqrt{2}\pi^{3/2}} \tag{5.80}$$

The overall error in the centroid—in the case of scintillation detectors with monolithic crystals and position-sensitive photodetectors—is the geometric sum of the three contributions from Equations (5.72), (5.76), and (5.80).

## REFERENCES

1. Lerche, C. W., J. M. Benlloch, F. Sánchez, N. Pavón, B. Escat, E.-N. Gimenez, M. Fernández, I. Torres, M. Giménez, A. Sebastià, and J. Cerdá. 2005. Depth of interaction within continuous crystals from the width of its scintillation light distribution. *IEEE Trans. Nucl. Sci.* 52:560–572.
2. Knoll, C. 2000. *Radiation detection and measurement*. 3rd ed., chapter 9. New York: Wiley & Sons.

3. Buzhan, P., B. Dolgoshein, L. Filatov, A. Ilyin, V. Kantzerov, V. Kaplin, A. Karakash, F. Kayumov, S. Klemin, E. Popova, and S. Smirnov. 2003. Silicon photomultiplier and its possible application. *Nucl. Instr. Meth. A* 504:48–52.
4. Otte, A., J. Barral, B. Dolgoshein, J. Hose, S. Klemin, E. Lorenz, R. Mirzoyan, E. Popova, and M. Teshima. 2005. A test of silicon photomultipliers as readout for PET. *Nucl. Instr. Meth. A* 545:705–715.
5. Herbert, D., V. Saveliev, N. Belcari, N. Ascenzo, A. Del Guerra, and A. Golovin. 2006. First results of scintillator readout with silicon photomultiplier. *IEEE Trans. Nucl. Sci.* 53:389–394.
6. Corsi, F., et al. 2006. Electrical characterization of silicon photo-multiplier detectors for optimal front-end design. *IEEE Nucl. Sci. Symp. Conf. Record* 2:1276–1280.
7. ORTEC Semiconductor Radiation Detectors. 2008. *Review of the physics of semiconductor detectors pulse formation process*. http://www.ortec-online.com/detectors/review_physics/pulse.htm.
8. Shockley, W. 1938. Currents to conductors induced by a moving point charge. *J. Appl. Phys.* 9:635.
9. Ramo, S. 1939. Currents induced by electron motion. *Proc. IRE* 27:584.
10. De Geronimo, G., P. O'Connor, V. Radeka, and B. Yu. 2001. Front-end electronics for imaging detectors. *Nucl. Instr. Meth. A* 471:192–199.
11. Corsi, F., M. Foresta, C. Marzocca, G. Matarrese, and A. Del Guerra. 2008. Experimental results from an analog front-end channel for silicon photomultiplier detectors. *IEEE Nucl. Sci. Symp. Conf. Record* 2:2010–2014.
12. Gray, P., P. Hurst, S. Lewis, and R. Meyer. 2000. *Analysis and design of analog integrated circuits*. 4th ed., chapter 11. New York: Wiley & Sons.
13. Radeka, V. 1988. Low noise techniques in detectors. *Ann. Rev. Nucl. Part. Sci.* 38:217–277.
14. Radeka, V. 1969. 1/f noise in physical measurements. *IEEE Trans. Nucl. Sci.* 16:17–35.
15. Bertuccio, G., A. Pullia, and G. De Geronimo. 1996. Criteria of choice of the front-end transistor for low-noise preamplification of detector signals at sub-microsecond shaping times for X- and gamma-ray spectroscopy. *Nucl. Instr. Meth. A* 380:301–307.
16. Wallmark, J. T. 1957. A new semiconductor photocell using lateral photoeffect. *Proc. IRE* 45:474–483.
17. Petersson, G. P., and L. E. Lindholm. 1978. Position sensitive light detectors with high linearity. *IEEE J. Solid-State Circuits* 13:392–399.
18. Emmons, R. B. 1967. The lateral photoeffect. *Solid-State Electr.* 10:505–506.
19. Schottky, W. 1930. Ueber den Entstehungsort der Photoelektronen in Kupfer-Kuperoxydul-Photozellen. *Physkal. Zeits.* 31:913–925.
20. Alberi, J. L., and V. Radeka. 1976. Position sensing by charge division. *IEEE Trans. Nucl. Sci.* 23:251–258.
21. Anger, H. O. 1958. Scintillation camera. *Rev. Sci. Inst.* 29, no. 1:27–33.
22. Anger, H. 1963. Gamma-ray and positron scintillation camera. *Nucleonics* 21, no. 10:56–59.
23. Anger, H. O. 1966. Sensitivity, resolution, and linearity of the scintillation camera. *IEEE Trans. Nucl. Sci.* 13:380–392.
24. Lauterjung, K., J. Pokar, B. Schimmer, and R. Stäudner. 1963. Grenzschichtzähler für Ortsund Energiebestimmung. *Nucl. Instr. Meth.* 22:117–121.
25. Charpak, G., J. Favier, and L. Massonnet. 1963. A new method for determining the position of a spark in a spark chamber by measurement of currents. *Nucl. Instr. Meth. A* 24:501–502.
26. McDicken, W. 1967. A position sensitive Geiger counter. *Nucl. Instr. Meth.* 54:157–158.

27. Lan, K. A., E. V. Hungerford, K. Johnson, D. Ambrose, P. Coffey, and J. J. Wang. 1995. Charge division read-out in a 1.8-m-long wire chamber. *IEEE Trans. Nucl. Sci.* 42:1430–1438.
28. Borkowski, C., and M. Kopp. 1970. Some applications and properties of one- and two-dimensional position-sensitive proportional counters. *IEEE Trans. Nucl. Sci.* 17:340–349.
29. Siegel, S., R. Silverman, Y. Shao, and S. Cherry. 1996. Simple charge division read-outs for imaging scintillator arrays using a multi-channel PMT. *IEEE Trans. Nucl. Sci.* 43:1634–1641.
30. Mukai, T. T., and M. Ishikawa. 1996. Resistive network for detecting the centroid of nonlinear coordinates. In *Industrial Electronics, Control, and Instrumentation, 1996: Proceedings of the 1996 IEEE IECON 22nd International Conference*. Vol. 2, 1052–1058.
31. Raffo, L., S. P. Sabatini, G. M. Bo, and G. M. Bisio. 1998. Analog VLSI circuits as physical structures for perception in early visual tasks. *IEEE Trans. Neural Networks* 9:1483–1494.
32. Hogan, T. K. 1943. A general experimental solution of Poisson's equation for two independent variables. *J. Inst. Eng. (Australia)* 15:89–92.
33. Hechtel, J. R., and J. A. Seeger. 1961. Accuracy and limitations of the resistor network used for solving Laplace's and Poisson's equations. *Proc. IRE* 49:933–940.
34. Vittoz, E. A. 1997. Pseudo-resistive networks and their applications to analog collective computation. In *Proceedings of the 7th International Conference on Artificial Neural Networks*, 1133–1150. London: Springer.
35. Radeka, V., and P. Rehak. 1979. Charge dividing mechanism on resistive electrode in position-sensitive detectors. *IEEE Trans. Nucl. Sci.* 26:225–238.
36. Masoliver, J., J. M. Porrà, and G. H. Weiss. 1993. Solution to the telegrapher's equation in the presence of reflecting and partly reflecting boundaries. *Phys. Rev.* E48:939–944.
37. Pullia, A., W. F. J. Muller, C. Boiano, and R. Bassini. 2002. Resistive or capacitive charge-division readout for position-sensitive detectors. *IEEE Trans. Nucl. Sci.* 49:3269–3277.
38. Davies, B. 2002. *Integral transforms and their applications*. New York: Springer.
39. Herrero, V., R. Colom, R. Gadea, C. W. Lerche, J. Cerda, A. Sebastia, and J. M. Benlloch. 2007. Front-end circuit for position sensitive silicon and vacuum tube photomultipliers with gain control and depth of interaction measurement. *Nucl. Instr. Meth. A* 576:118–122.
40. Tornai, M. P., L. R. MacDonald, C. S. Levin, S. Siegel, and E. J. Hoffman. 1996. Design considerations and performance of a 1.2 $cm^2$ beta imaging intra-operative probe. *IEEE Trans. Nucl. Sci.* 43:2326–2335.
41. Popov, V., S. Majewski, and B. Welch. 2006. A novel readout concept for multianode photomultiplier tubes with pad matrix anode layout. *Nucl. Instr. Meth. A* 567:319–322.
42. Olcott, P., J. Talcott, C. Levin, F. Habte, and A. Foudray. 2005. Compact readout electronics for position sensitive photomultiplier tubes. *IEEE Trans. Nucl. Sci.* 52:21–27.
43. Seidel, J., J. Vaquero, F. Barbosa, I. Lee, C. Cuevas, and M. Green. 2000. Scintillator identification and performance characteristics of LSO and GSO PSPMT detector modules combined through common X and Y resistive dividers. *IEEE Trans. Nucl. Sci.* 47:1640–1645.
44. Zhang, N., C. Thompson, D. Togane, F. Cayouette, and K. Nguyen. 2002. Anode position and last dynode timing circuits for dual-layer BGO scintillator with PS-PMT based modular PET detectors. *IEEE Trans. Nucl. Sci.* 49:2203–2207.

45. Lerche, C., M. Döring, A. Ros, V. Herrero, R. Gadea, R. Aliaga, J. Colom, F. Mateo, J. Monzó, N. Ferrando, J. Toledo, J. Martínez, A. Sebastiá, F. Sanchez, and J. Benlloch. 2009. Depth of interaction detection for gamma ray imaging. *Nucl. Instr. Meth. A* 600:624–634.
46. Rogers, J., D. Saylor, R. Harrop, X. Yao, C. Leitao, and B. Pate. 1986. Design of an efficient position sensitive gamma ray detector for nuclear medicine. *Phys. Med. Biol.* 31:1061–1090.
47. Matthews, K. L., II, S. M. Leonard, C. E. Ordonez, D. E. Persyk, and W. Chang. 2001. A depth-encoding Anger detector using scintillation fibers. *IEEE Trans. Nucl. Sci.* 48:1397–1402.
48. Landi, G. 2002. Properties of the center of gravity as an algorithm for position measurement. *Nucl. Instr. Meth. A* 485:698–719.

# 6 Design Considerations for Positron Emission Tomography (PET) Scanners Dedicated to Small-Animal Imaging

*Rejean Fontaine*

## CONTENTS



## 6.1 INTRODUCTION

Positron emission tomography (PET) aims to quantitatively measure the distribution of a radiotracer injected into a subject. The injected tracer is generally chosen to target bioprocesses of interest occurring in specific cells or tissues.[1] Among tracers, the

most widely used in PET is indisputably the fluorodeoxyglucose (FDG) for cancer studies and diagnosis.[2] Currently, researchers focus on new tracers aimed at better understanding human gene properties or disorders and support development of new drugs.[3] The imaging research under this perspective is better known as molecular imaging, where animal models play a crucial role.[4] According to the results of a 2003 survey,[5] discoveries in this research field are so important that molecular imaging has been considered among the 10 technologies that will most affect human health. Since 2000, many small-animal PET scanners have been proposed to answer genomic and proteomic requirements.[6,7] In contrast with clinical PET, small-animal imaging faces two major challenges originating from (a) the animal model's small internal organ sizes and from (b) geometric side effects caused by the scanner's small diameter. Many issues such as spatial resolution, sensitivity, and high count rates, along with the ability to support multimodality imaging, depict the modern challenges where trade-offs are constantly and carefully considered.

In this race to acquire the clearest image at the lowest cost, the detector and electronic front end are among the important factors limiting the ideal performances of PET scanners. Research teams or companies able to handle the many aspects of the scanner (detectors, electronics, real-time signal processing, computer engineering, nuclear physics, nuclear medicine, and end users such as biologists or applied research physicians) generally better succeed at building performing systems. Even if medical research is the utmost goal of scanner design, this chapter will focus on the detector and sensitive electronics required to meet medical imaging expectations. The first section will give a physics overview to understand how images can be obtained from PET scanners; the second section will focus on scanner design trade-offs chosen to meet imaging performance requirements; and the third will present front-end electronics associated with various detectors. As the current state of the art is mostly focused on application-specific integrated circuit (ASIC) design and performance, the discussion of electronics will be oriented in this way. The chapter concludes with a discussion of multimodality challenges.

## 6.2 PHYSICS OF PET SCANNERS

Many of the trade-offs taken when building a small-animal PET scanner come from physics phenomena such as Compton scattering, electronic noise, and the characteristics of the selected photodetectors. Even if some phenomena present negligible effects, their combination makes PET troublesome to master. An understanding of these phenomena helps in maximizing the scanner performance.

A PET image acquisition is initiated with the injection of a tracer in a subject. After a certain amount of time, the tracer reaches the targeted cells or tissues where it will be metabolized. In PET, the tracer—called a *radiotracer*—is composed of a molecule such as glucose, protein, peptide, etc., in which one of its constituents has been substituted with a radioisotope ($^{18}$F, $^{15}$O, $^{13}$N, $^{11}$C). The PET camera reports quantitatively the distribution of the radioisotopes carried by the tracer. The radioisotopes disintegrate themselves with a half-life decay time specific to the radioactive atom ($^{18}$F ≈ 110 min, $^{15}$O ≈ 2 min, $^{13}$N ≈ 10 min, and $^{11}$C ≈ 20 min).[8] In the case of PET, the radioisotope nucleus ejects a neutrino ($\nu$), some energy, and a positron ($\beta^+$),

an antimatter particle. The positron, which is the only part of interest in PET, loses its kinetic energy in collisions with neighboring atoms until it annihilates with a surrounding electron. The mean distance traveled before the positron annihilation is a function of the $\beta^+$ energy at its ejection from the nucleus ($^{18}F \approx 0.5$ mm, $^{15}O \approx 0.4$ mm, $^{11}N \approx 0.5$ mm). This displacement generates an uncertainty in the location of the disintegration and thus defines a physical limitation in spatial resolution for PET.

The annihilation generates two 511-keV photons emitted collinearly with an uncertainty of ±0.25° (Figure 6.1). The annihilation photons are generally stopped by scintillating crystal-based detectors (Figure 6.2) spread on a camera ring (Figure 6.3). The scintillators are made of heavy materials such as lutetium, gadolinium, or

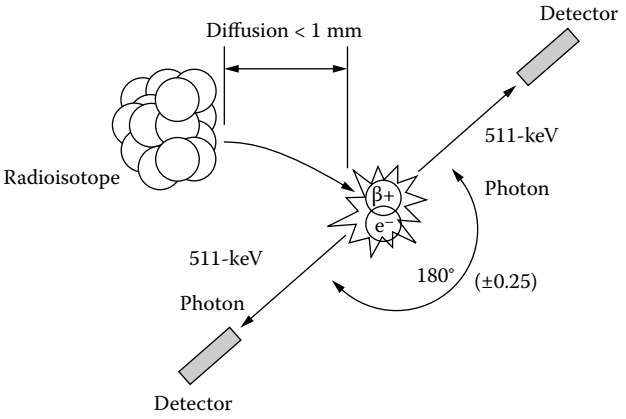



**FIGURE 6.1** The disintegration of a radioisotope generates a positron that annihilates with an electron in the environment to give two 511-keV photons emitted collinearly.

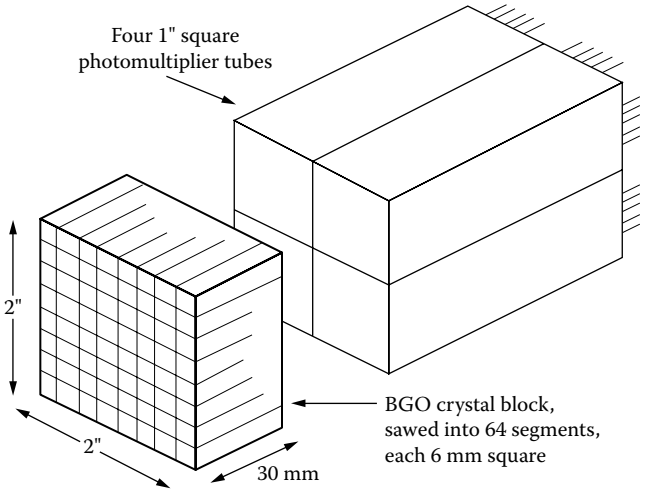



**FIGURE 6.2** A BGO crystal array coupled to a pack of four PMTs is used to stop 511-keV photons and generate an electric pulse needed to extract relevant information in PET.

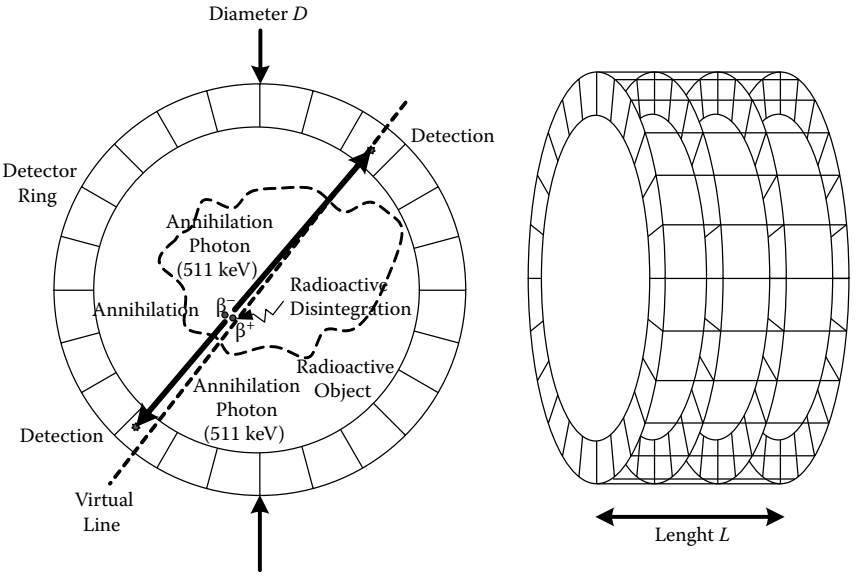



**FIGURE 6.3** A ring of detectors of diameter $D$ and axial length $L$ intercepts the annihilation photons. Detection of two photons within a determined timing window creates a coincidence pair.

bismuth (Figure 6.2). The crystal's characteristics influence the performance of the scanner in terms of timing and energy resolution, which, in turn, affect the contrast and the spatial resolution.

The camera ring, with a diameter $D$ and an axial length $L$, tries to match intercepted photons—called *events* or *singles*—in pairs using three conditions: a coincidence timing window (0.5 to 20 ns), a bounded energy range (200 to 650 keV), and a field of view within the detector ring (usually 60 degrees from each detector's perspective with the opposite side) (Figures 6.2 and 6.3). Matched event pairs are called *coincidences*.

Three different cases can occur when measuring the singles (Figure 6.4). The first one is seen when two 511-keV photons issued from the same annihilation deposit all their energy as a photoelectric effect in a scintillating crystal of the camera ring (Figure 6.4a). The two individual photons, called *singles*, create a true coincident event represented as a line of response (LOR). This forms the ideal case and represents the desired data required for image reconstruction.

In the second case, the two singles detected are not coming from the same disintegration. One must understand that annihilation is isotropic, and only photons traveling toward the scanner ring can be intercepted. Other photons simply escape. If only one of the two annihilation photons of two distinct disintegrations reaches the camera ring, a false coincidence—called *random coincidence*—is detected (Figure 6.4b). Unlike singles in true coincidences that are time correlated because they are issued from the same annihilation, photons in random coincidence pairs have no timing relationship. This means that the probability of detecting a random

coincidence is the same for any time difference, whether it is 1, 10, or 100 ns. So, the random rate is uniform regardless of the timing window width, as can be seen on the edges of a timing-resolution spectrum (Figure 6.5), where the timing difference of events picked up by two detectors is charted on a histogram. What is seen in this figure is the sum of the trues (the Gaussian shape) and the randoms (flat floor). The problem is that trues and randoms are indistinguishable using the detectors or the three criteria. The whole detected coincidences are thus referred to as *prompts*. The timing resolution, which corresponds to the width of the Gaussian shape, is generally

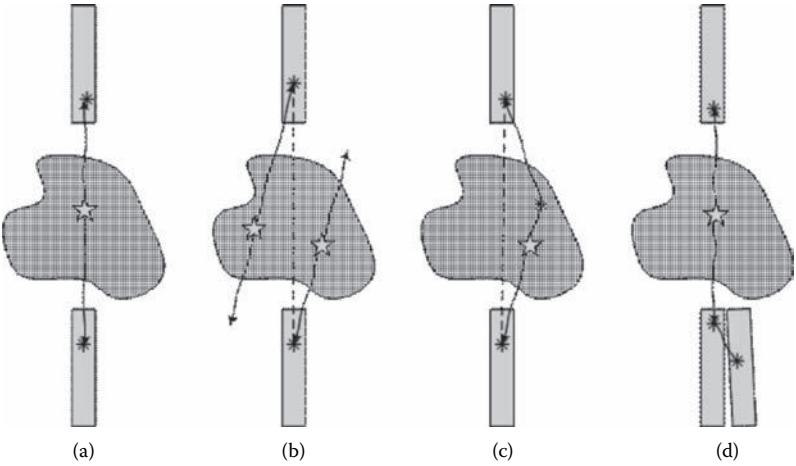



**FIGURE 6.4** Three types of coincidence: (a) true coincidence, (b) random coincidence, (c) Compton scatter in the subject, and (d) Compton scatter in the crystal detector.

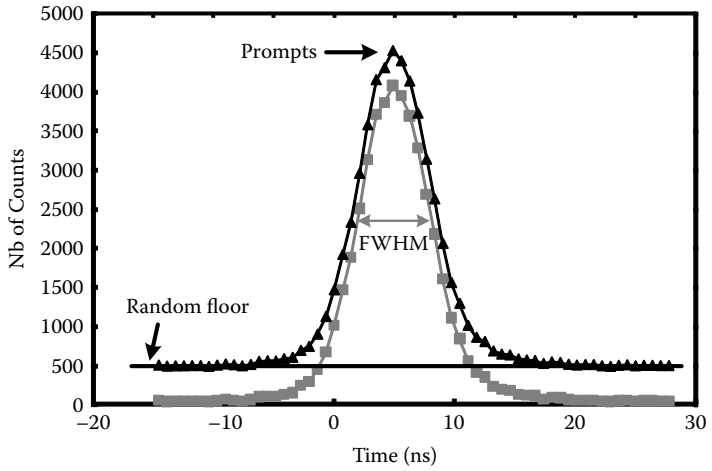



**FIGURE 6.5** The timing-resolution spectrum is built by charting the timing difference of coincidence event pairs in a histogram. The spectrum contains two sections: the Gaussian-like shape issued from true coincidence and a flat floor issued from the random coincidence.

expressed as a full width at half maximum (FWHM)* value in units of nanoseconds. The timing resolution is a factor of the detector and the electronic noise, as will be discussed in the next section. Random coincidences can be statistically estimated, but it is preferable to narrow as much as possible the timing window to reduce their number at the source. Reducing the timing window will also reduce the count rate in the scanner and relax the requirements of the output data links.

In the third case, an annihilation photon produces a Compton scatter within the subject or within the crystal detector (Figure 6.4c,d). The original annihilation photon then deviates from its original trajectory with an angle θ, leaving a part of its energy to the environment (Figure 6.6). This phenomenon can be described by the Klein-Nishina equation[9]

$$\lambda' - \lambda = \frac{h}{m_e c}(1 - \cos\theta) \tag{6.1}$$

where λ and λ′ are the wavelengths before and after the scattering, $h$ is Planck's constant, $m_e$ is the mass of the electron, and $c$ is the speed of light.

When occurring inside the subject, Compton scatter generates a false LOR (Figure 6.4c). An energy threshold is often applied to limit the coincidences to singles without Compton effects. The Compton scattering can be observed on an energy spectrum, where the measured energy of individual events is charted on a histogram (Figure 6.7). Such a spectrum in scientific papers contains, most of the time, four sections. The first one is the photopeak, which contains the events that make photoelectric interactions. The Gaussian shape is centered at 511 keV. Since a Compton scatter is not possible over 350 keV, a valley exists between the photopeak and the Compton events. At low energy, the electronic noise is added to the Compton scatter, which makes a negative slope on this part of the spectrum. Unlike timing resolution, the energy resolution is calculated in percent by computing $\Delta E/E$, which corresponds to evaluating the FWHM divided by the mean of the Gaussian. Energy resolution depends not only on the detector used, but also on the system noise. For this reason, many papers will add a fourth component to the spectrum, corresponding to fake events generated with a pulser. In this case, the pulser energy is selected to be higher than the photopeak and its Gaussian width (in FWHM) is only related to the noise in the system. Using this information, one can estimate how the system degrades performance, and one can also deconvolve the system noise from the measurement to obtain the energy resolution associated to the detector itself with

$$\sigma_{measure} = \sqrt{\sigma^2_{\text{noise detector}} + \sigma^2_{\text{noise system}}} \tag{6.2}$$

This section presented the main physical phenomena related to PET. The basics of timing resolutions and energy resolutions were also discussed in an effort to better understand how to avoid undesirable coincidences in images. The next section will focus on a higher level, which is the PET scanner itself. Many trade-offs related to

* FWHM is an equivalent measurement of the standard deviation. FWHM is ≈2.35 SD.

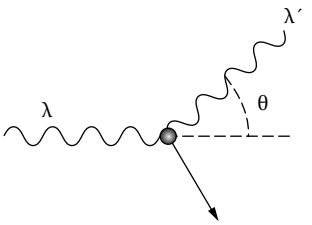



**FIGURE 6.6** The Compton scatter effect.

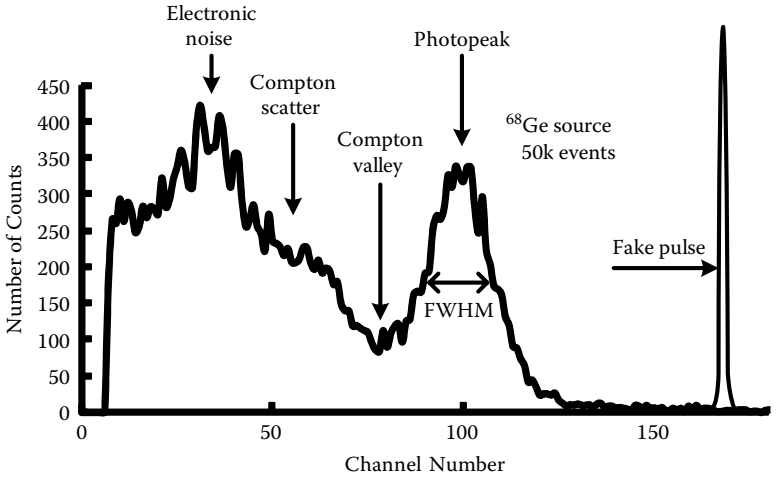



**FIGURE 6.7** The energy resolution is built by charting the energy of individual events in a histogram. The spectrum contains the photopeak, the Compton valley, the Compton scatter, the electronic noise, and occasionally a pulser peak.

the physics covered in this section will be highlighted, with the focus on how these are related to scanner performances.

## 6.3 TRADE-OFFS IN SCANNER DESIGNS

In addition to physical phenomena, the design of a scanner must also deal with many trade-offs to maximize image resolution and sensitivity or minimize cost. Current scanners can hardly manage all of these trade-offs at the same time, and choices must be made that will affect final performance or to minimize costs.

One of the main advantages of PET over other imaging modalities is its ability to quantitatively measure bioprocesses down to picomolar concentrations. This measurement is obtained with the reconstruction of an image composed of pixels depicting the distribution of the radiotracer within a subject slice. The contrast in the image provides information about radiotracer's concentration. As small-animal organs are, indeed, "small," pixels must be scaled accordingly. However, the pixel size cannot

be indefinitely reduced. Spatial resolution is defined in PET as the smallest voxel* containing quantitative information, set by[9]

$$FHMW = a\sqrt{\left(d/2\right)^2 + b^2 + (0.0022D)^2 + r^2} \tag{6.3}$$

where $a$ is a factor related to the image reconstruction algorithm ($1.1 < a < 1.3$); $d$ is the detector size (i.e., the crystal in our case); $b$ is a coding factor, 0.0022, the uncertainty of the collinearity of the disintegration; $D$ is the diameter of the scanner; and $r$ is the mean positron displacement before annihilation. $D$ and $r$ are physical constraints one must live with, the scanner diameter sized to the animal to be studied, and the positron displacement associated with the radiotracer that must be used for the experiment. However, the crystal size $d$ and the coding factor $b$ can play an important role for spatial resolution, and judicious choices can be made to maximize scanner performance.

As seen in Equation (6.3), the smaller the crystal size is, the better the theoretical spatial resolution will be. On the other hand, this parameter can cause other problems, as will be seen later, and cannot be infinitely reduced. Moreover, cost is an issue, as thin crystals are hard to manufacture. Another issue is that PET is limited by the positrons' diffusion. The spatial resolution of scanners will always be bound by the latter factor, and it is not worth designing detectors with intrinsic spatial resolution better than the positron diffusion.

The coding factor $b$ is a consequence of the detector type used. Many scanners rely on a block detector composed of a photomultiplier tube (PMT) coupled to a pixelated matrix of crystals (Figure 6.2).[10] A decoding scheme similar to Anger logic is used to calculate the center of gravity of the interaction. The Anger logic scheme consists of measuring the ratio among the four PMT outputs (W, X, Y, Z).[1] The ratio determines in which "individual" crystal the interaction occurred, and the sum of the PMT signals corresponds to the energy deposited in the crystal. The presence of Compton scatter as well as the uncertainty in the localization of the interacting photon increases the $b$ factor, which leads to a degradation in the reconstructed image. More recently, some research teams have proposed coupling crystals individually to a light detector such as an avalanche photodiode (APD).[11–12] Such an approach reduces the $b$ factor to almost zero at the cost of an increase in electronic channels and in system architecture complexity.

The small size of rat and mouse organs, together with the arrival of new radiotracers that are very specific to certain cells or tissues, require the measurement of very low radioactivity concentrations. The scanner must then be able to intercept as many annihilation photons as possible in a reasonable amount of time to create a useful image. This parameter, called *scanner sensitivity*, is estimated by the number of detected true coincidences divided by the number of expected disintegrations. The number of true coincidences is a factor of the solid angle seen by a source localized in the middle of the scanner.[1] A longer scanner will cover a higher solid angle and thus possess a higher sensitivity, but it will require a larger budget to purchase the

* A voxel is a three-dimensional pixel.

extra detectors. The crystal length can be another factor, but it is usually chosen to stop >95% of annihilation photons. However, short crystals could be used in PET/MRI (PET/magnetic resonance imaging) systems where the PET rings are inserted inside an MRI tube. In this case, the limited maneuvering space margin inside the MRI enforces the use of short crystals to leave enough room for the mouse to be scanned, to the detriment of sensitivity.

Crystal size also plays a crucial role in sensitivity through the packing fraction. A highly pixelated matrix of crystals usually contains crystals individually covered with reflecting foils that produce voids in the matrix[13,14] where photons can slip through. This phenomenon is also present between neighboring crystal matrices. The ideal scanner would be a crystal tube with detector matrices, which is costly and hard to manufacture. However, many scanners tried to maximize the solid angle by employing large monolithic crystals, where a decoding scheme is used to the detriment of the spatial resolution (i.e., high decoding factor *b*).[15]

Highly pixelated crystals may also affect scanner sensitivity through losses caused by Compton scatter. In practical scintillating crystals (i.e., lutetium oxyorthosilicate, LSO), only ≈30% of first interactions produce a photoelectric effect; the remaining ≈70% generate a Compton scatter. As the crystal size reduces, the probability of stopping the second interaction within the same crystal gets smaller. One must understand that two Compton scatters within the same crystal are indistinguishable, and the energy measured will be the sum of the two scatters. The performance loss can appear in two ways, depending on the crystal detector used. When a block detector is used, the total energy is detected, but a decoding error will occur that will reduce the spatial resolution. In the same conditions, when individual crystals are coupled to an APD, both Compton scatters found in nearby channels can be under the energy threshold and thus be discarded, which will reduce sensitivity. To circumvent these aforementioned limitations, research projects are attempting to reduce the energy threshold and localize the Compton scatters in a pixelated matrix to introduce them in the image along with the true coincidences.[16]

The sensitivity could also be affected by dead time, a phenomenon where events near in time are lost for different reasons. Dead time follows a mix of two behaviors: paralyzable and nonparalyzable.[9] Paralyzable dead time occurs when piled-up events are detected and the system cannot accept a new event until the dead time associated with the last event in the pileup is completed. A pileup event occurs when the rate of events is higher than the recovery time, i.e., the time for the electronic front end to get back to a steady state. The paralyzable dead time is associated, most of the time, with the detector and the accompanying electronics. For example, if a simple threshold is used and a second photon interacts in the crystal before the signal comes back under the threshold, the system will only see the first photon. Large detectors are more sensitive to paralyzable dead time. On the other hand, nonparalyzable dead time is associated with data processing. One can see nonparalyzable dead time as a first-in–first-out buffer that does not accept any more events when the buffer is full. The system can process one event, but if two events are received, one of them is discarded regardless of its occurrence relative to other events.

To meet spatial resolution performance, the crystal detectors used in small-animal PET scanners are usually long, thin, and pointed toward the center of the field of view (FOV) (Figures 6.2 and 6.4). Since all lines of response are positioned by default at the crystal tip located in the center of the scanner, the non-null size of the crystals generates what is called a *parallax effect*. When the disintegration arises in the periphery of the scanner, the depth of interaction (DOI) of the annihilation photon in the crystal exacerbates the uncertainty about the localization of the interaction. This effect creates a disparity in the spatial resolution within the field of view, where the spatial resolution is maximal in the center of the scanner and degrades radially. Although there are many techniques available to mitigate DOI, all of these methods are expensive, and DOI is often relegated as a secondary concern in scanner designs. Among these methods, three are of interest for their trade-offs between resolution, complexity, and cost.

In the first scheme, the DOI is estimated with the varying pulse decay time issued from a crystal having a gradient of cerium concentration along the crystal.[17] When coupled to an appropriate signal processor, the DOI can be approximated with a resolution <2 mm. However, this technique is not popular, since it is still hard to produce such crystals.

The second scheme measures the difference of light intensity on both ends of the crystal tips,[18,19] It is accomplished by using two detector sets. Commonly, one detector is a PMT and the other is an APD. The APD is used on the tip pointing toward the scanner to ensure a good solid angle—PMT being large and long compared to APD. This technique is not cost effective because it uses twice the electronics compared to other solutions. However, resolutions ≈1.5 mm can be reached with this scheme. The same solution can be used in the axial direction by using DOI measurements to determine the axial position. Many very long crystals are placed side by side, where both crystal ends are read out with a PMT (Figure 6.8).[20] In this scheme, the axial localization is determined by the ratio of the energy measured at both tips of the crystal, while

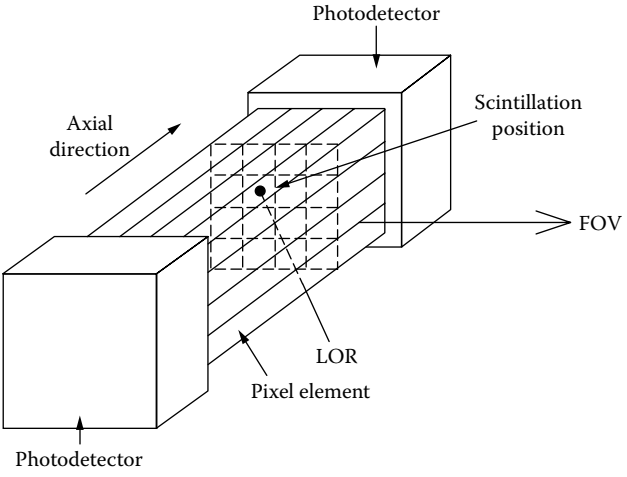



**FIGURE 6.8** Photodetectors at each end of the crystal matrix perform a DOI measurement to localize the position of the interaction along the crystal in the axial direction.

the radial DOI is determined by the crystal that has scintillated in the matrix. Good spatial resolution of ≈1.5 mm can also be obtained using this approach.

The final scheme uses a stack of crystals with different scintillating characteristics named *phoswich*.[21,22] A light detector such as an APD or a PMT reads the light emitted by the stack, and signal processing identifies the scintillating crystal. This approach requires only one detector at the crystal end and, when coupled with adequate signal processing, obtains high performance in the localization of the scintillating crystal. However, the DOI is limited to the crystal length used.

Scanner design involves many trade-offs to accommodate the physics effects discussed here. Choices made on the mechanical architecture of detectors as well as the type of detector affect the scanner performance. The next section will give more details about the electronics required to maximize scanner performance.

## 6.4 TYPICAL ELECTRONICS CHAIN FOR PET

Many scanners have been designed with the ultimate goal of localizing the positron annihilation. Different techniques, like wire chambers[23] or crystal scintillator-based systems,[24,25] present their own pros and cons for manufacturability/cost and ease of use. The scintillator-based system is the most widespread, and the next few subsections will focus on this particular system.

A PET electronic chain is essentially a light-to-voltage amplifier. The basic principle of a scintillator-based system consists of transforming the energy of an annihilation photon into a bunch of visible photons required to generate electrons in a photosensitive detector. Normally, the shape of the electron pulse follows an exponential shape with a decay time of $\tau$ (Figure 6.9). A transimpedance amplifier, or a charge-sensitive preamplifier (CSP), transforms the electron pulse into a voltage pulse, depending on the quantity of electrons issued from the detector, the CSP being a very low noise amplifier. The voltage pulse can be shaped/filtered to maximize signal-to-noise ratio and improve timing and energy measurements. Prior to these measurements, the signal is DC-adjusted in a baseline restoration circuit to maximize the performance of the other electronic stages.

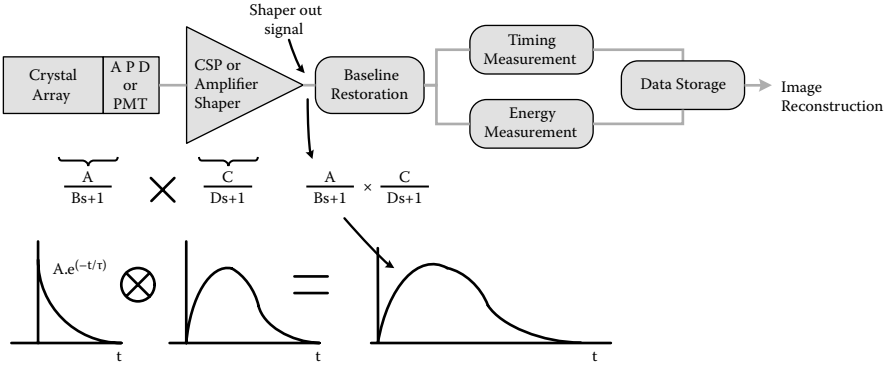



**FIGURE 6.9** Typical data-acquisition chain for a PET scanner.

The timing measurement can be relative or absolute. In the case of relative timing measurement, the event trigger opens a timing window and looks on the opposite side of the scanner to see if there is another event within the determined time. If so, digital data for both channels, such as detector addresses, are sent to the data storage, regardless of the event type. The first scanners used this approach for its lower complexity, although to the detriment of scanner flexibility. In the case of absolute timing, individual events receive a digital time stamp directly at the detector. This time stamp, together with the energy measurement and the address, is sent to a coincidence engine that parses the data and sends the coincidence events to a data storage unit for subsequent image reconstruction. This approach has the highest flexibility but requires complex electronics and an early digitization of information.

PET images rely on two principal measurements: the energy of the photon and the time of occurrence. Time-of-occurrence measurements are based on evaluating when a signal crosses a voltage threshold (Figure 6.10). The noise in the signal affects the timing resolution through the signal slope, where the uncertainty on the signal amplitude is projected on the timing axis. Figure 6.10 presents signals with the same noise amplitude with different slopes, showing that the steeper the slope at the threshold crossing, the better is the timing resolution.

This phenomenon can be expressed mathematically as

$$\sigma_t = \frac{\sigma_{\text{noise}}}{\delta V / \delta t} \tag{6.4}$$

where the numerator is the quadratic sum of all noncorrelated noise (described in Section 6.4.2.3), and the denominator is the slope of the signal at the crossing of the threshold. The designer's objective will be both to reduce the amount of noise and to increase the slope at the threshold crossing.

The energy measurement is made by integrating the number of the collected charges and measuring the maximum on the shaper-out signal (Figure 6.9). If the integration constant is long enough compared to the crystal's decay time, there is no ballistic deficit, and the voltage obtained at the output of the integrator is proportional to the energy deposited in the crystal. However, noise is still present at this maximum point, and uncertainty on the energy measured will decrease the energy resolution.

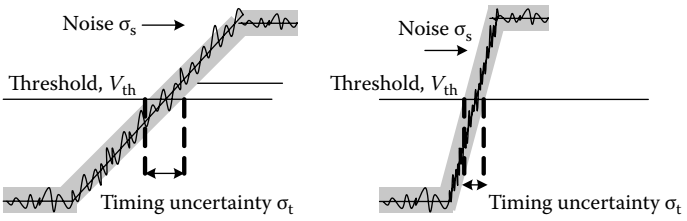



**FIGURE 6.10** The effect of the slope on timing resolution in the presence of noise. A steeper slope represents a lower uncertainty for the same noise variance.

The next subsections will describe each stage of a typical electronic chain for PET in greater detail, how to extract relevant information from annihilation events as depicted in Figure 6.9, and what the major contributors to noise are.

### 6.4.1 The Crystal

Scintillating crystals stop annihilation photons and transform their energy into light. The NaI (made in 1948)[26] crystal has been the reference for many decades, but it has slowly been replaced by the cerium-doped lutetium oxyorthosilicate (LSO) since the early 2000s. LSO has a high stopping power at 511 keV, a fast decay time, and similar light output to NaI. Many crystals are derived from the original LSO recipe, where different impurities such as gadolinium, yttrium, etc., give different optical properties to the scintillating crystal. Table 6.1 presents the key properties of common scintillating crystals.[27]

The shape of light emission in a scintillating crystal follows an exponential decay time ($A \cdot e^{-t/\tau}$) highly contaminated with noise. The variance ($\sigma_s$) of this noise comes from two uncorrelated noise sources, $\sigma_e$ and $\sigma_i$[28]

$$\sigma_s = \sqrt{\sigma_e^2 + \sigma_i^2} \tag{6.5}$$

where $\sigma_e$ is a statistical variance with a Poisson distribution, and $\sigma_i$ is the intrinsic crystal resolution. The quantity $\sigma_e$ can be expressed as

$$\sigma_e = 2.35\sqrt{F_s / \overline{N}} \tag{6.6}$$

where $F_s$ is the excess noise due to statistical fluctuations in APD gain (see Equation [6.11]), and $\overline{N}$ is the effective mean number of primary photoelectrons created in

**TABLE 6.1**
**Generic Properties of Common Scintillators**

| | LGSO | LYSO | LSO | BGO |
|---|---|---|---|---|
| Decay time (ns) | 65–75 | 40 | 40 | 40–300 |
| Fast/slow ratio | | | | (1/10) |
| Light output (APD) [a] | 45 | 85 | 75 | 30 |
| Light output (PMT) [a] | — | 75 | 75 | 15 |
| Peak emission (nm) | ≈415 | 420 | 420 | 480 |
| Index of refraction | ≈1.8 | 1.81 | 1.82 | 2.15 |
| Density (g/cm³) | 6.5 | 7.19 | 7.35 | 7.13 |
| Absorption coefficient at 511 keV (1/μm) | 14.3 | 12.6 | 12.3 | 11.6 |
| Effective Z | 58 | 63 | 64 | 73 |
| Probability of PE (%) | ≈30 | ≈28 | 34 | 44 |

[a] Relative to NaI(Tl).

the APD. The intrinsic crystal resolution $\sigma_i$ originates from inhomogeneities in the crystal and other variations in the light collection, such as crystal defects, inhomogeneous surface coating, photon traps, and depth of interactions of the annihilation photon. The total combined noise elements highly distort the temporal distribution of photons, and the pure exponential decay time as depicted in Figure 6.9 will never be seen. Rather, the waveforms always slightly differ from one another. For this reason, a shaper is used to integrate the signal and smooth/filter the noise as much as possible without deteriorating the signal-to-noise ratio at the frequency of interest. All these noise sources affect timing and energy measurements in different ways, as will be discussed later.

Photon emission decay time in crystals plays an important role in scanner performance, as can be seen in Figure 6.9. In the Laplace transform domain, the shaper-out signal can be seen as the multiplication of the crystal's impulse response and the shaper's impulse response, which is equivalent to a convolution in the time domain. A fast decay time produces shaper-out signals with steeper slopes (i.e., includes higher frequencies), which allows better timing measurement, as seen in Equation (6.4). Moreover, a fast decay time generates shorter shaper-out pulses, which enable higher count rates and may help to increase sensitivity. As shown in Table 6.1, common decay times for different crystal species vary from 600 to 300 ns for BGO. New crystal breeds such as Luteium ortholuminate can now obtain decay times down to 17 ns, but these are not always suited to be coupled with all photodetectors such as APD. Finally, the decay time can be modified with the introduction of some impurities such as cerium in some crystal breeds since emission wavelength may not be compatible.

Decay time is not the only factor contributing to scanner performance; the light output, expressed in units of photons per MeV (Ph/MeV), is also an important parameter. The first photons—called the *primary photons*—emitted by the crystal contain most of the timing information. Having a high number of primary photons generates a shaper-out signal with higher amplitude and thus a steeper slope at the threshold crossing. In this case, the number of photons represents the exponential's integral and not its amplitude at $t_{0+}$. For this reason, the literature presents the amplitude of the exponential as

$$p = \frac{(\text{Ph/MeV}) * \text{photon energy}}{\tau} * e^{-t/\tau} \tag{6.7}$$

where the photon energy is 511 keV in the case of PET, and $\tau$ is the decay time of the crystal. The reader can perform the exponential integration to verify that the total number of photons is correct.

Matching a crystal with a photodetector must be done carefully because the emission wavelength of the crystal and the absorption wavelength of the photodetector may be different. For this point, the APD is particularly selective compared to PMT. Table 6.1 presents peak emission wavelengths for different crystal species, where LSO (≈420 nm) is well suited for APD. A bad crystal coupling regarding the wavelength would produce only a few electrons at the output of the photodetector and is equivalent to a low-light output.

Many crystals presented in the literature are hygroscopic, e.g., NaI. These crystals must be hermetically sealed in an enclosure to avoid water absorption, which makes the crystal literally melt. LSO-like crystals are very attractive because they are not hygroscopic and because of their fast decay time and light output property.

Crystal matrix assembly is a science in itself. One simple way involves cutting trenches in a monolithic block, as illustrated in Figure 6.2. The depth of the trenches is carefully controlled to minimize optical cross talk between pixelated crystals while ensuring a good light collection. This assembly is used for block detection, where the position is decoded with an Anger logic scheme. For highly pixelated photo detectors like APD, the assembly of individual crystals is preferred. In this case, the surfaces of individual crystals receive a special treatment to reflect, collect, or diffuse light. The available surface treatments include mechanical polish, chemical polish, and usage of diffusing/reflecting foils between crystals.[29,30] A coupling material such as an optical glue or grease with adequate refraction index is also used to maximize the light transfer between the crystal and the photodetector.

### 6.4.2 Common Detectors in PET

The PMT is indisputably the most widely used photodetector in PET. However, the new challenges associated with small-animal imaging compel the designers to look at new photodetectors. Recently, the APD has been used in commercial scanners, and the scientific community is very enthusiastic about the silicon photomultiplier (SiPM). The following subsections describe these photodetectors and their associated electronics in greater detail.

#### 6.4.2.1 The Photomultiplier Tube and Its Amplification Stage

The PMT is a vacuum tube sensitive to light, with a detection efficiency of ≈40%. It is composed of a photocathode, several dynodes, and an anode housed in a glass envelope (Figure 6.11). The incident photons strike the photocathode material, where electrons are emitted by photoelectric effect. Electrons are accelerated toward the first dynode by a high-voltage electrical field. By the time they reach this dynode,

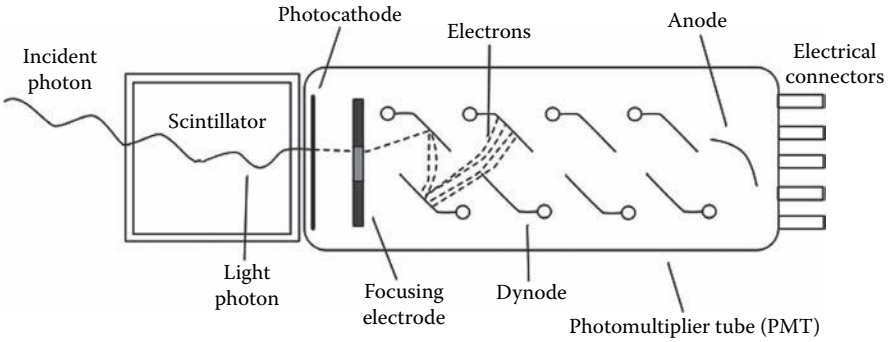



**FIGURE 6.11** A photomultiplier tube is composed of a cathode, multiple dynodes, and an anode housed in a glass envelope.

they have accumulated sufficient energy to generate secondary low-energy electrons, which in turn are accelerated to the subsequent dynode stages. The anode converts the amplified electron surge into an electric-pulse signal wired toward subsequent signal-processing stages. The PMT geometry is designed to create a cascade of electrons and enable amplification factors of up to $10^6$–$10^7$ with a high signal-to-noise ratio. These characteristics make the PMT a detector of choice for clinical PET imaging. In fact, most clinical and a few small-animal PET scanners are based on PMT detector blocks, as shown in Figure 6.2.

Initial detector blocks evaluated the interaction position through an Anger logic scheme, but modern PMT with multiple anodes (called *position-sensitive PMT*, or PSPMT)[31] simplify the decoding, as the charge distribution can be directly determined and read out without complex circuitry. As PMTs have excellent signal-to-noise ratios, the electronics used to transform the output signal is straightforward and usually made of cascaded, simple operational amplifiers (OPAMP) in transimpedance mode. The only OPAMP characteristic requirement is a band-pass frequency wide enough to meet adequate rising edges, thereby achieving good timing resolutions. Some PMTs can achieve timing resolution ≈160 ps, a characteristic suited for clinical time-of-flight (TOF) PET.[32]

Other decoding schemes were proposed to localize the annihilation-photon interaction within block detectors. Among them, the artificial neural network (ANN)[33,34] is of interest because it could replace the matrix by a continuous crystal, which reduces the manufacturing costs of the detector at the expense of a higher complexity in the back-end electronics.

The block-detector approach engenders inexpensive scanners, since many "optical" channels (i.e., the individual crystal pixels) share four electronic amplifiers (the PMTs) with minimal and low-cost supporting electronics. Therefore, back-end electronics sections such as constant-fraction discriminators (CFD) are also shared. Moreover, as the signal-to-noise ratio (SNR) is very high, excellent timing resolutions (<1 ns) can be obtained. However, as discussed in Section 6.3, the spatial resolution suffers from the decoding scheme, and small crystals of $1.5 \times 1.5$ mm produce the same spatial resolution as crystals of $2 \times 2$ mm with direct crystal–APD coupling. Furthermore, because pixels are multiplexed, block detectors suffer from pileup problems at high count rates, unlike individually coupled detectors (i.e., APD). Finally, the relatively large size of PMT is not well suited to the high density electronics of small-animal scanner rings.

Partly because of their better detection efficiencies, new detectors such as APD and SiPM should slowly replace PMT. But the major blow comes from multimodality scanner-design considerations, specifically for PET/MRI, where PMTs cannot be used because they contain magnetic materials and are too large for small animal scanners. And rather than making a PET system tailored for MRI, it is more cost effective to design one architecture that fits them all.

#### 6.4.2.2 Avalanche Photodiode-Based Detectors and Their Amplification Stage

The avalanche photodiode (APD) is a silicon-based detector. It is composed of a p–n junction operated in reverse voltage. Roughly, an APD has an absorption region, where photons generate carriers (electrons), and a high-voltage bias multiplication region.[35] Gains of 100, and more recently 200,[36] obtained from APD contrast with PMT gains of four to five orders of magnitude greater. Unlike PMT, the APD has a good detection efficiency (≈70%), which makes it more sensitive to photon detection. APDs are insensitive to magnetic fields and can be easily pixelated with sizes down to the square millimeter ($mm^2$), which makes them suitable to be coupled to individual pixel crystals. However, the low gain of APD combined with the noise generated with the reverse p–n junction makes them difficult to work with. The APD's performance is limited by its leakage current and the shot noise coming from the presence of a reverse voltage. The shot noise in the p–i–n (PIN) diode is a statistical Poisson fluctuation of the dark current, described by

$$i_n = \sqrt{2qBI_D} \tag{6.8}$$

where $q$ is the electronic charge, $B$ is the system bandwidth, and $I_D$ is the dark current fluctuation. In APD, the bulk leakage current ($I_{DB}$) is unfortunately multiplied by the gain $M$.

$$I_D = I_{DS} + I_{DB}M \tag{6.9}$$

then

$$i_n = \sqrt{2q(I_{DS} + I_{DB} \cdot M^2 \cdot F) \cdot B} \tag{6.10}$$

where $F$ is the excess noise factor coming from the statistical nature of the avalanche[38] and is computed as

$$F = Mk_{eff} + \left(2 - \frac{1}{M}\right)(1 - k_{eff}) \tag{6.11}$$

where $k_{eff}$ is the effective ionization ratio. The multiplication of the bulk leakage current with the gain makes the APD a noisy detector. Combined with the detector's low gain, APD signals have little amplitude with a relatively poor signal-to-noise ratio compared to PMT. These problems make APD-based systems hard to design, and special care must be taken within the first electronic stages. Advanced noise-control techniques applied to APD systems are described in the next subsection.

#### 6.4.2.3 Noise Issues for APD-Based Detectors

In APD-based systems, the measurements are all affected by the presence of noise, which can be seen as a local variation of the amplitude (in voltage) of the signal. Even

if noise amplitude is bounded, it is always expressed in term of spectral density. As the different noise contributions are not correlated, they can be square-root summed, as shown in Equation (6.12)

$$\sigma_{noise}(t) = \sqrt{\sigma_s^2(t) + \sigma_{photostatistic}^2(t) + \sigma_{electronic}^2(t)} \tag{6.12}$$

where $\sigma_s(t)$ is calculated using Equation (6.5)

$$\sigma_{photostatistic}^2(t) = qFM^2 \int_{-\infty}^{+\infty} I_{photo}(\alpha)\omega^2(t-\alpha)d\alpha \tag{6.13}$$

and

$$\sigma_{electronic}(t) = \sigma_{ENC} \tag{6.14}$$

where $I_{photo}$ is the current induced by the scintillation light,[37] $\omega(t)$ is the weighting function extracted from the shaper filter,[38] and $\sigma_{ENC}$ is the equivalent noise charge from the electronic front end.[40] The photostatistic variance, $\sigma^2_{photostatistic}(t)$, as presented in Equation (6.13), is reported before the APD (Figure 6.9) and is defined by the convolution of the current generated with the light emitted from the crystal with an exponential shape, as in Equation (6.7), and the weighting function to the square. Indeed, the weighting function is determined by the normalized impulse response of the shaper.

From Equation (6.13), one can see that the noise variance is time varying. This means that the noise changes in time and obtains its maximum value at the maximum of the weighting function. The shaping filter type and order play an important role in timing resolution, as they shape the slope of the shaper-out signal and influence the quantity of photostatistic noise at the threshold crossing (Figure 6.10 and Equation [6.4]). Faster shapers enable shaper-out signals with higher slopes. However, if the choice of the integration constant is too small compared to the crystal decay time, there is a risk of a ballistic deficit[9] for the energy measurement. On the other hand, the integration constant must be too long to introduce undesired dead time in the scanner. Good timing resolution reduces the number of random events, and good energy resolution reduces the Compton scatter. This clearly demonstrates the trade-off between the timing resolution and the energy resolution.

Unlike the photostatistic noise, the equivalent noise charge (ENC) is time invariant. It can be seen as fluctuations of carriers in the electronic front end, whose parts are CMOS (complementary metal oxide semiconductor) technology. These carriers inject uncorrelated time pulses. Moreover, the ENC noise is permanently present in the shaper-out signal, in contrast to the photostatistic noise, which is only present when a pulse is present. The ENC noise has three uncorrelated components called parallel noise, serial noise, and 1/*f* noise. The parallel noise, also called *shot noise*, represents the fluctuations in the number of carriers emitted in the APD. As the APD is a reverse p–n junction, the carriers flow in one direction caused by the applied electric field and present noise pulses at the input of the CSP. The parallel noise is modeled with a

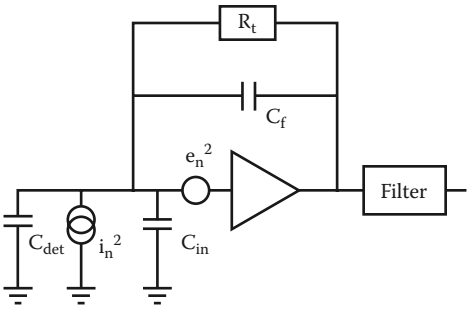



**FIGURE 6.12** The noise model for an APD-based detector consists of a current source in parallel with the APD detector and a voltage source in series with the CSP.

current source $i_n^2$ in parallel with the APD (Figure 6.12). The spectral density of the parallel noise is significant at low frequency and can be computed with

$$ENC_p = \sqrt{ENC_{APD}^2 + \left(\frac{1}{q}\right)^2 (qI_{leak-feedback} + 2kT\alpha_{n-feedback} g_{m_{feedback}})\tau_{peak} A_p} \qquad (6.15)$$

where $ENC^2_{APD}$ is the equivalent noise from the APD, $I_{leak}$ is the leakage current on the feedback circuit in the CSP, $\alpha_{n-feedback}$ is a parameter depending on the transistor type and its operating region, $g_m$ is the transconductance seen from the feedback circuit, $\tau_{peak}$ is the peaking time of the shaper filter, and $A_p$ is a parameter related to the kind of filter used. These parameters and equations are well covered in the literature.[39–40] One important note is that the parallel noise is not affected by the input transistor, unlike other noise sources. Usually, the APD noise dominates the parallel noise.

The serial noise, also called *thermal noise*, is dominated by the thermal processes in the resistors created by the conducting channels of the electronic front end. The CSP topology must be carefully chosen to ensure that the input transistor noise dominates over the noise from subsequent amplification stages. The folded cascade topology is a candidate of choice in this case.[41] Unlike parallel noise, the serial noise is not unidirectional, and noise fluctuations can be positive and negative. Serial noise spectral density is negligible at low frequency and more pronounced at high frequency. A voltage source $e_n^2$ in series with the CSP input models both the serial noise and the flicker noise described in the next paragraph. The serial noise can be estimated with

$$ENC_s = \frac{1}{q}\sqrt{\frac{2kT\alpha_{n-input}}{g_{m_{input}}}(C_{det} + C_{in})^2 \frac{A_s}{\tau_{peak}}} \qquad (6.16)$$

where $C_{det}$ is the detector capacitance, $C_{in}$ is the input capacitance of the CSP, and $A_s$ is a parameter related to the shaper filter used. One can notice that serial noise increases with the size of the input transistor through $C_{in}$ and decreases with the

transconductance of the input transistor. As $g_m$ differs depending on whether the transistor is in strong, moderate, or weak inversion, the size of the input transistor must also introduce this parameter in the whole equation.

Finally, the third noise source, called the *flicker-noise* (or 1/*f* noise), also comes from the input transistor and is also modeled with a voltage source $e_n^2$ in series with the input of the CSP. The flicker noise appears as charge fluctuations in the gate–channel interface, which influences the conduction in the transistor channel. The flicker noise can be estimated with

$$ENC_{1/f} = \frac{1}{q}\sqrt{\pi \frac{K_f}{C_{ox}WL}(C_{det} + C_{in})^2 Af} \tag{6.17}$$

where $K_f$ is the 1/*f* noise coefficient of the input transistor (technology dependent); *W* and *L* are the width and length of the transistor channel, respectively; and $A_f$ is a parameter related to the shaper filter used. The flicker noise is dependent on the transistor type (pMOS or nMOS), with nMOS transistors having more 1/*f* noise than pMOS. One can notice that the bigger the transistor is (i.e., *W* and *L* are large), the smaller the flicker noise will be. This fact contrasts with the minimization of the serial noise, and a trade-off must be made to minimize the overall noise of the electronic front end.

As noted previously, the ENC noise is frequency dependent. A better way to present this information consists in plotting the ENC versus the peaking time of the shaper filter (Figure 6.13). As the shaper function also affects the photostatistic noise through the slope of the shaper-out signal, it is easier to find the best solution to timing and ENC using Figure 6.13.

Unlike PMT, APD-based detectors are hard to design. However, APD systems offer the possibility to individually couple a pixel crystal, a photodetector, and an electronic chain thanks to highly integrated CMOS circuits, thereby reducing the

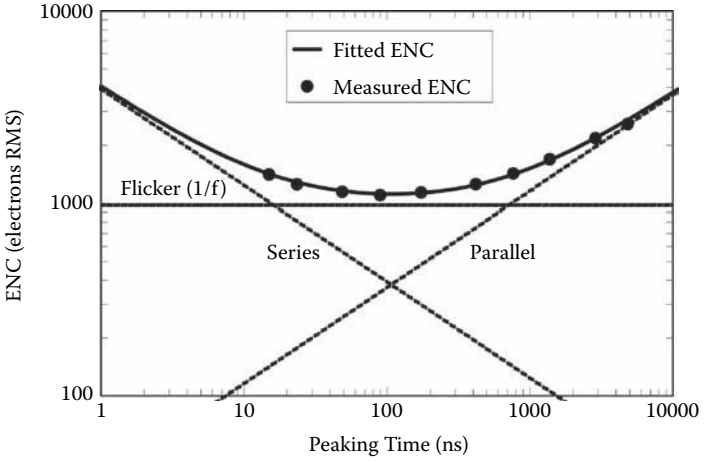



**FIGURE 6.13** The total ENC noise is the square-root sum of the parallel, serial, and 1/*f* noise.

decoding factor $b$ to zero. This approach maximizes the spatial resolution at the expense of more complex electronics.

#### 6.4.2.4 Silicon Photomultipliers (SiPMs)

Silicon photomultipliers are actually the latest trends in the scientific community. They consist of APD detectors tied in parallel and operated in Geiger mode. They present the same advantage of APD (size, nonmagnetic materials) but with a gain varying from $10^5$ to $10^7$, like PMT. SiPMs of limited size are currently available and present excellent timing resolution (<1 ns). The operation mode is quite simple; the junction is biased above the breakdown voltage. When a photon hits the sensitive region, an electron-hole pair is generated. If the high electric field is strong enough, both the electrons and holes generate avalanche effects. Upon detection of the avalanche, a passive or active quenching circuit reduces the biasing under the breakdown voltage to stop the avalanche process before the cell's destruction. The number of generated charges is equal to the cell capacitance multiplied by the overvoltage (i.e., biasing voltage minus breakdown voltage).

$$Q = C(V_{bias} - V_{Breakdown}) \tag{6.18}$$

The output signal is proportional to the number of cells that fired. For this reason, SiPM presents drawbacks. The number of cells must be selected according to the number of light photons that may hit the SiPM. There must be more cells than expected photons; otherwise, a signal saturation occurs, and the energy measurement becomes nonlinear. For a certain number of photons to be detected, the cell size cannot be infinitely reduced, as a dead area surrounds every cell. To evaluate this phenomenon, the notion of photodetection efficiency (PDE) has been created. The PDE is expressed as

$$PDE = QE \cdot \in \cdot P_{trigger} \tag{6.19}$$

where QE is the quantum efficiency, $\in$ is the ratio of the sensitive area versus the total area, and $P_{\text{trigger}}$ is the probability that an incoming photon triggers a breakdown. The number of cells that may fire can then be expressed as

$$N_{firedcells} = N_{Total}\left(1 - e^{N_{Photon} \cdot PDE / N_{Total}}\right) \tag{6.20}$$

A second set of drawbacks to SiPM includes the presence of dark count rate and cross talk. The dark count rate is issued from carriers trapped in the silicon. These carriers are released for several microseconds (μs) after the pulse signal. Optical cross talk occurs when a photon crosses the cell boundary and triggers an adjacent one. An optical isolation is mandatory to avoid this problem. These drawbacks limit the recovery time of SiPM and significantly affect the overall count rate performance of the detector.

Unlike APD, SiPM does not need complex front-end electronics. A simple transimpedance amplifier is sufficient to transform the charge pulse into a voltage pulse. However, the remaining back end is still very similar to APD design, as individual SiPM channels may be coupled to an individual electronic channel. The electronics complexity then remains very similar to that of APD systems.

#### 6.4.2.5 Baseline Restoration and Holding

The shape-out signal usually contains a DC component originating from electronic bias or from leakage current present in the detector. As these circuits are usually built in ASIC, an intermediate stage is required to adjust the DC level to optimize timing and energy-measurement performance in subsequent electronic stages. At first glance, a simple adjustment of the DC baseline would seem sufficient to correct the situation. However, pulse pileup can affect the DC level of the shaper-out signal. To counter this case, an analog memory is used to maintain the DC level even in presence of pileup at high count rates. Two approaches are available: the baseline restorer and the baseline holder.

The former approach uses a simple capacitor as analog memory to maintain the DC level. In this scheme, the DC component of the signal without pulse is maintained in a capacitor and presented at the input of a comparator used to trigger the time of occurrence.[42] When a pulse is detected, a switch removes the signal from the comparator input and leaves only the DC component stored in the capacitor. This practice requires high-frequency circuitry that can generate noise in the surrounding electronics. The latter approach uses a feedback circuit that compares the shaper-out signal to a voltage reference ($V_{\text{baseline}}$). A low-pass filter with a very low frequency generates a current signal at the input of the CSP[43] or at the input of the shaper[44] through a transconductance circuit (Figure 6.14). Between these two solutions, the second is preferred because the transconductance can be seen as noise in parallel with the detector or the shaper input. In the second solution, the impact on the signal-to-noise ratio is less important, as the pulse signal is already amplified by the CSP. The feedback approach requires only low-frequency components, which minimizes the power consumption of large systems and ensures a stable baseline to the timing comparator, which is required to trigger the time of occurrence in the next stage.

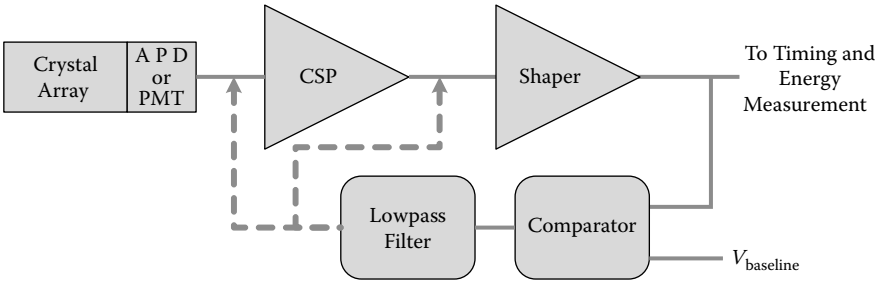



**FIGURE 6.14** The baseline holder uses a feedback circuit.

### 6.4.2.6 Energy and Timing Measurement

Energy measurement with PET electronics is usually carried out by sampling the pulse's maximum amplitude as seen at the shaper's output (Figure 6.15a). This method requires an analog-to-digital converter (ADC), which is usually unsuitable for large systems because it requires a large area and considerable power consumption. A second method uses time-over-threshold (TOT), which is a measure of the pulse's time spent over a fixed threshold.[45] This ADC-free approach requires high-resolution time-to-digital converters and a conversion table, because the measurement is not linear.

Energy measurement performance is tied to the type of shaper used and to the noise present at the shaper-out's sampling point, i.e., the maximum amplitude. Shapers may be of different type and order: CR-RC$^n$, semi-Gaussian, bipolar, or monopolar.[9,46] Depending on the design objective, either bipolar or monopolar filters can be used. Bipolar filters offer the advantage of fluctuating around a stable DC level. In this case, the timing information is triggered at the crossing of the stable DC level (Figure 6.16a). However, the bipolar filter requires more time to return to a stable state compared to a monopolar filter,[9] which may affect the scanner dead time.

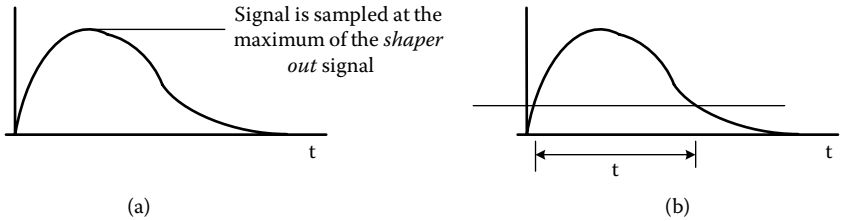



**FIGURE 6.15** The energy-measurement examples with (a) pulse amplitude and (b) time-over-threshold.

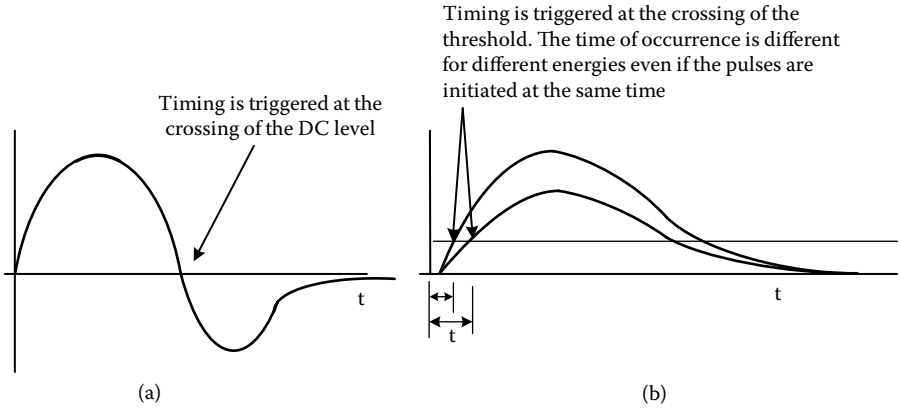



**FIGURE 6.16** Pulses with different shapes produced by (a) a bipolar shaper and (b) a monopolar shaper.

Applying a simple threshold for timing extraction with monopolar filters is not desirable, as timing is affected by the signal's energy (Figure 6.16b, different crossing points for two simultaneous events). Two main options are used to correct the energy variation with monopolar filters. The first one consists of measuring the energy with a TOT and then correcting the time of occurrence accordingly. This circuit requires the minimum electronics, as the same time-to-digital converter can be used for both the timing and the energy measurement. The second method relies on a constant-fraction discriminator (CFD).[9,47] A CFD generates a fraction of the shaper-out signal and subtracts it from a delayed version. The resulting bipolar signal crosses the DC level independently of the original signal's amplitude. However, CFD delay lines are hard to implement on ASIC, and one CFD can serve only a few electronic channels without generating dead time in the scanner. For these reasons, CFDs are not used on scanners with complex electronics like APD-based systems.

## 6.5 MULTIMODALITY

Up to now, PET/CT scanners have demonstrated their superiority in obtaining better diagnostics compared to single-modality scanners.[48] For this reason, other multimodal scanners were proposed starting in 2000, where the PET scanner was coupled to a morphological scanner such as MRI or computed tomography (CT). The many combinations of PET, SPECT, DOT, MRI, and CT were explored, but PET/MRI and PET/CT scanners are the most widely used. PET/MRI is valued for its ability to image soft tissues (like the brain) and to minimize the X-ray dose injected into the subject, and PET/CT is preferred for high speed imaging, i.e., heart gating.

Designing and building a multimodal scanner is an order of magnitude more challenging than constructing monomodal scanners, as new requirements emerge from the new apparatus. For example, the PMT detectors commonly used in PET system cannot be used in PET/MRI systems, as they are large and contain magnetic materials. New detectors based on APD are being designed for this purpose. Another example can be found in PET/CT systems applied to molecular imaging. This field requires ultra-low-dose X-ray imaging to minimize interaction with the bioprocess under scrutiny. Conventional CT modality, based on the integration of the photon flux, induces a significant dose and must be rethought. In this case, the photon-counting approach of PET can be advantageously used to reduce the X-ray dose without compromise on the image contrast but with lower image resolution compared to conventional CT. Photon-counting-based X-ray uses the same electronic chain to acquire both the PET photon (511 keV) and X-ray photon (30–120 keV). These systems reach a high electronic complexity, which is inherent for detectors individually coupled to an electronic chain. The multimodality will now pave the way to the next generation of PET.

## 6.6 CONCLUSION

This chapter presented the challenges related to scanner design, including physical phenomena and detector selection, along with accompanying electronics. Choices made in the early stages of the design process greatly influence the final performance

of the scanner. Many trade-offs must be closely considered when matching the electronic front end with the detector, depending on its type and its features. These decisions greatly influence the scanner architecture, functionality, and flexibility. Beyond front-end electronics, other questions remain. Should the electronic circuits include real-time computation or push the workload off-line? Is the scanner a portable or a fixed machine? How can calibration be automated for biasing, energy threshold adjustment, or estimation of random coincidence? These kinds of questions may inflate the complexity of the overall architecture. One must not forget that small-animal scanners are built to help the scientific community in biology, genomics, and pharmacology to understand bioprocesses at the base of life. Operating a scanner is not a straightforward task, and everything must be done to simplify calibration and maintenance tasks for the end users. This is usually where the commercialization of such equipment succeeds or fails.

Designing a PET scanner requires the management of many trade-offs involving particle physics, nuclear medicine, electronics, as well as electrical engineering. The main challenge involves managing a multidisciplinary team, where the contribution of every member is of the utmost importance.

## REFERENCES


1. Cherry, S., J. Sorenson, and Michael Phelps. 2003. *Physics in nuclear medicine*. 3rd ed. Philadelphia: Saunders/Elsevier.
2. Cook, G. J., et al. 2004. Pitfalls and artifacts in $^{18}$FDG PET and PET/CT oncologic imaging. *Sem. Nucl. Med*. 34 (2): 122–133.
3. Lecomte, R., et al. 2002. Tomographie d'émission par positrons en oncologie. *Phys. Can.* 58 (2): 109–117.
4. Massoud, T. F., and S. S. Gambhir. 2003. Molecular imaging in living subjects: Seeing fundamental biological processes in a new light. *Genes Dev*. 17:545–580.
5. Huang, G. T. 2003. 10 emerging technologies that will change your world: Molecular imaging. *Tech. Rev*. 106:42.
6. Townsend, D.W., et al. 2003. PET/CT scanners: A hardware approach to image fusion. *Sem. Nucl. Med*. 33:193–204.
7. Goertzen, A. L., et al. 2002. Simultaneous molecular/anatomical imaging of mouse in vivo. *Phys. Med. Biol*. 47:4315–4328.
8. Tétrault, M. A., et al. 2008. System Integration of the LabPET™ small animal PET scanner. *IEEE Trans. Nucl. Sci*. 55:2546–2550.
9. Knoll, G. 2000. *Radiation detection and measurement*. 3rd ed. New York: Wiley.
10. Derenzo, S. E., and W. W. Moses 1993. Critical instrumentation issues for resolution <2 mm, high sensitivity brain PET. In *Quantification of brain function, tracer kinetics and image analysis in brain PET*, ed. K. Uemura, N. A. Lassen, T. Jones, and I. Kanno, 25–40. New York: Elsevier.
11. Casey, M. E., and R. Nutt. 1986. A multicrystal two-dimensional BGO detector system for positron emission tomography. *IEEE Trans. Nucl. Sci*. 33:460–463.
12. Vaska, P., et al. 2004. RatCAP: Miniaturized head-mounted PET for conscious rodent brain imaging. *IEEE Trans. Nucl. Sci*. 51:2718–2722.
13. Bergeron, M., et al. 2009. Performance evaluation of the LabPET™ APD-based digital PET scanner. *IEEE Trans. Nucl. Sci*. 56:10–16.

14. Spanoudaki, V., et al. 2008. Effect of temperature on the performance of proportional APD-based modules for gamma ray detection in positron emission tomography. *IEEE Trans. Nucl. Sci*. 55:469–840.
15. Pichler, B. J. 2001. A 4×8 APD array, consisting of two monolithic silicon wafers, coupled to a 32-channel LSO matrix for high-resolution PET. *IEEE Trans. Nucl. Sci*. 48:1391–1396.
16. Vandenbroucke, A., and C. S. Levin. 2008. Study of scintillation crystal array parameters for an advanced PET scanner dedicated to breast cancer imaging. *Proc. Nucl. Sci. Symp*. 4914–4919.
17. Maas, M. C., et al. 2006. Experimental characterization of monolithic-crystal small animal PET detectors read out by APD arrays. *IEEE Trans. Nucl. Sci*. 53:1071–1077.
18. Michaud, J. B., et al. 2009. Monte Carlo results from neural networks as an alternative to Compton photons LOR analysis. *Proc. Nucl. Sci. Symp*. Orlando, Florida, October 2009.
19. Inadama, N., et al. 2002. A depth of interaction detector for PET with GSO crystals doped with different amounts of Ce. *IEEE Trans. Nucl. Sci*. 49:629–633.
20. Miyaoka, R. S., T. K. Lewellen, H. Yu, and D. L. McDaniel. 1998. Design of a DOI PET detector module. *IEEE Trans. Nucl. Sci*. 45:1069–1073.
21. Moses, W. W., and S. E. Derenzo. 1994. Design studies for a PET detector module using a pin photodiode to measure depth of interaction. *IEEE Trans. Nucl. Sci*. 41:1441–1445.
22. Salvador, S., et al. 2009. Design of a high performances small animal PET system with axial oriented crystal and DOI capability. *IEEE Trans. Nucl. Sci*. 56:17–23.
23. Lecomte, R., et al. 2001. Performance analysis of phoswich/APD detectors and low-noise CMOS preamplifiers for high-resolution PET systems. *IEEE Trans. Nucl. Sci*. 48:650–655.
24. Ziemons, K., et al. 2006. The ClearPET™ LSO/LuYAP phoswich scanner: A high performance small animal PET system. *IEEE Trans. Nucl. Sci*. 53:25–29.
25. Schafers, K. P., et al. 2005. Performance evaluation of the 32-module quadHIDAC small-animal PET scanner. *J. Nucl. Med*. 46:996–1004.
26. Tai, Y.-C., et al. 2003. MicroPET II: Design, development and initial performance of an improved microPET scanner for small-animal imaging. *Phys. Med. Biol*. 48:1519–1537.
27. Spanoudaki, V., D. P. McElroy, and S. I. Ziegler. 2006. An analog signal processing ASIC for a small animal LSO-APD PET tomograph. *Nucl. Instrum. Meth. Phys. Res. A* 564:451–462.
28. Hofstadter, R. 1948. Alkali halide scintillation counters. *Phys. Rev*. 74:100–101.
29. Pepin, C. M., et al. 2009. Depth-of-interaction (DOI) with fast scintillators read out by APDs using LabPET digital crystal identification algorithms. Paper presented at 10th International Conference on Inorganic Scintillators and Their Applications, Jeju, Korea, June 2009.
30. Lecomte, R., et al. 1998. Investigation of GSO, LSO and YSO scintillators using reverse avalanche photodiodes. *IEEE Trans. Nucl. Sci*. 45:478–482.
31. Carrier, C., and R. Lecomte. 1990. Theoretical modelling of light transport in rectangular parallelepipedic scintillators. *Nucl. Instrum. Meth. Phys. Res. A* 292:685–692.
32. Strul, D., J. Sutcliffe-Goulden, P. Halstead, and P. K. Marsden. 2002. Optimization of fiber-optic readout of LSO scintillation crystals with acid etching. *IEEE Trans. Nucl. Sci*. 49:619–623.
33. Hamamatsu Global. 2010. http://www.hamamatsu.com/.
34. Kyba, C. C. M., et al. 2008. Energy and timing response of six prototype scintillators for TOF-PET. *IEEE Trans. Nucl. Sci*. 55:1404–1408.
35. Aliga, R. J., et al. 2006. Corrected position estimation in PET detector modules with multi-anode PMTs using neural networks. *IEEE Trans. Nucl. Sci*. 53:776–783.

36. Bruyndonckx, P. 2004. Neural network-based position estimators for PET detectors using monolithic LSO blocks. *IEEE Trans. Nucl. Sci*. 51:2520–2525.
37. PerkinElmer. 2006. Application note: Avalanche photodiode, a user guide. www.optoelectronics.perkinelmer.com.
38. Bérard, P., et al. 2008. LabPET II, a novel 64-channel APD-based PET detector module with individual pixel readout achieving submillimetric spatial resolution. *Proc. IEEE Nucl. Sci. Symp*. 5457–5462.
39. Casey, M., et al. 2003. Analysis of timing performance for an APD-LSO scintillation detector. *Nucl. Instrum. Meth. Phys. Res. A* 504:143–148.
40. Radeka, V. 1988. Low-noise techniques in detectors. *Ann. Rev. Nucl. Part. Sci*. 38:217–277.
41. Lecomte, R., et al. 2001. Performance analysis of Phosswich/APD detectors and low-noise CMOS preamplifiers for high-resolution PET systems. *IEEE Trans. Nucl. Sci*. 48:650–655.
42. O'Connor, P., and G. De Geronimo. 2002. Prospects for charge sensitive amplifiers in scaled CMOS. *Nucl. Instrum. Meth. Phys. Res. A* 480:713–725.
43. De Geronimo, G., and P. O'Connor. 2005. MOSFET optimization in deep submicron technology for charge amplifiers. *IEEE Trans. Nucl. Sci*. 52:3223–3232.
44. Pratte, J. F., et al. 2004. Front-end electronic for the RatCAP mobile animal PET scanner. *IEEE Trans. Nucl. Sci*. 51:1318–1323.
45. East, L. V. 1970. Baseline stabilization in direct coupled counting systems. *Rev. Sci. Instr*. 41:1245–1246.
46. Fiorini, C., and T. Frizzi. 2006. Baseline holder for CMOS readout circuits with feedback extension to the charge preamplifier. *Proc. IEEE Nucl. Sci. Symp*. 406–409.
47. De Geronimo, G., P. O'Connor, and J. Grosholz. 2000. A CMOS baseline holder (BLH) for readout ASICS. *IEEE Trans. Nucl. Sci*. 47:818–822.
48. Manfredi, P. F., et al. 2000. Noise limits in a front-end system based on time-over-threshold signal processing. *Nucl. Instrum. Meth. Phys. Res. A* 439:361–367.
49. Okawa, S., M. Yoshizawa, and K. Husimi. 1976. Direct synthesis of the Gaussian filter for nuclear pulse amplifiers. *Nucl. Instr. Meth*. 138:850–892.
50. Binkley, D. M., et al. 2002. A 10-Mc/s, 0.5-µm CMOS constant-fraction discriminator having built-in pulse tail cancellation. *IEEE Trans. Nucl. Sci*. 49:1130–1140.
51. Charron, M., et al. 2000. Image analysis in patients with cancer studied with a combined PET and CT scanner. *Clin. Nucl. Med*. 25:905–910.

# 7 Geiger-Mode Avalanche Photodiodes for PET/MRI

*Jae Sung Lee and Seong Jong Hong*

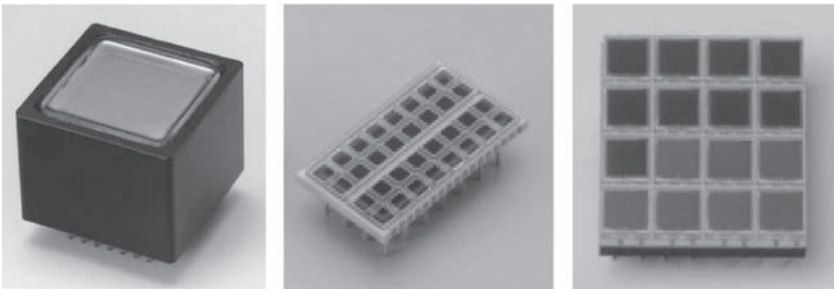
## CONTENTS

## 7.1 MOTIVATION

Nuclear medicine imaging systems, like positron emission tomography (PET) and single-photon-emission computed tomography (SPECT), rely on the use of scintillation detectors, which consist of inorganic scintillation crystals, and a photomultiplier tube (PMT). Visible light emitted by a scintillation crystal, due to its interaction with incident radiation, is converted into an electrical signal by a photocathode. This signal is then amplified by a strong electric field between the cascading electrodes of a PMT, which is the most common type of photodetector used in the high-energy physics and medical imaging fields, and has been for more than 70 years (Figure 7.1a). However, PMTs are expensive, because many of their mechanical components are handmade. Furthermore, in the case of combined MRI and PET or SPECT units, PMTs must be intensively shielded from magnetic fields.

The avalanche photodiode (APD) is a semiconductor that also provides signal amplification following the interaction between incident radiation and a scintillation crystal, though in this case the amplification is due to the a cascade of electron-hole pairs (Figure 7.1b). APD detectors have attracted the attentions of those developing fully integrated PET/MRI systems because they are insensitive to the magnetic fields generated by MR units, and because they are substantially smaller than PMTs, which is an important practical consideration. However, current APD-based PET/MRI systems are limited by a relatively low gain (ca. 1000), which requires the incorporation of a low-noise preamplification system, which increases PET detector volume and slow response time. Accordingly, much research effort has been directed at developing photodetectors with much higher gains and faster responses than those possible with current APDs.

The recently developed Geiger-mode APDs (G-APDs) are solid-state detectors with almost the same gain as PMTs (Figure 7.1c). G-APDs are composed of an array of micro APDs operated in Geiger mode at bias voltages above the breakdown voltage. Although individual Geiger-mode microcells do not provide a proportional output, they do provide a signal proportional to the energy of incident radiation when configured in parallel. Furthermore, because G-APDs operate in Geiger mode rather



(a) (b) (c)

**FIGURE 7.1** Photodetectors: (a) photomultiplier tube (PMT), (b) conventional avalanche photodiode (APD), and (c) Geiger-mode APD (G-APD) (© Hamamatsu Photonics. With permission.)

than in proportional mode, they provide more gain than conventional APDs. They are also relatively compact and cheap. Therefore, G-APD technology is viewed favorably by those developing integrated PET/MRI and SPECT/MRI systems. In this chapter, we introduce the approaches used to develop G-APD PET detector systems for use in PET/MRI units, and we compare PMT and APD technologies.

## 7.2 INTRODUCTION TO PET/MRI

### 7.2.1 PET AND PET/CT

#### 7.2.1.1 PET

Modern medical imaging methods can be categorized as anatomic or functional modalities. Anatomic modalities include those that image morphological abnormalities, such as X-ray computed tomography (CT) and magnetic resonance imaging (MRI). On the other hand, functional modalities image physiologic, functional, or biochemical characteristics, such as cerebral blood flow, oxygen and glucose metabolism, and the binding of ligands to receptors. Nuclear medicine imaging (or radioisotope imaging) is currently the most widely used type of functional imaging technology. Furthermore, PET is the most advanced nuclear medicine imaging technology, and has many physical and biochemical advantages over other functional imaging modalities, such as planar scintigraphy and SPECT.

Prior to PET scans, a small amount of radioisotope or radio-pharmaceutical (a pharmaceutical bonded to a radioisotope) is usually injected into a vein. The injected radio-pharmaceutical (tracer) then targets and is localized at regions where the biochemical process of interest is undertaken. At a set time after injecting the tracer, a PET scan is initiated, and emitted gamma rays are detected. The data so obtained is then processed to produce tracer distribution images.

The nuclei of the radioisotopes used for PET scans are intrinsically unstable, and they decay by simultaneously ejecting a positron (a positively charged antimatter equivalent of an electron) and neutrino to form a stable nucleus. Emitted positrons lose kinetic energy due to interactions with surrounding matter and come to rest within $\approx 10^{-9}$ s, whereupon they interact destructively with electrons in a high-energy process that results in mutual annihilation and the emission of two 511-keV gamma rays in opposing directions.

PET scanners detect these gamma ray pairs and provide spatial and temporal information on the distribution of the positron-emitting tracer. The most commonly used gamma ray detection device is the scintillation detector, in which pixelated scintillation crystals, such as BGO, LSO, or LYSO, are coupled with photodetectors. To detect emitted gamma rays, these crystals have higher densities and contain elements with higher atomic masses than the NaI(Tl) crystals used in gamma cameras. The electronic components in PET scanners analyze scintillation "events," and determine whether gamma-ray pairs are detected simultaneously using a coincidence circuit. This circuit rejects gamma ray pairs that are detected with a time interval longer than a certain threshold (the timing window), which is usually set from a few hundred picoseconds to several nanoseconds.

#### 7.2.1.2 PET/CT

By combining PET with anatomical imaging modalities like CT or MRI, one can synergistically enhance the clinical information obtained; for example, accurate anatomic localizations, correlation studies, partial volume correction, and statistical reconstruction are possible. Before the development of integrated multimodal imaging modalities, only software fusion was possible (e.g., Lee et al. 2005), and although software fusion technology is used in brain imaging research, its clinical usefulness is limited because it is time consuming and user unfriendly. More importantly, software fusion technology is often less than satisfactory when applied extracranially because of movement artifacts during the time delays between scans using different modalities.

On the other hand, PET/CT units use an integrated hardware system in a single gantry (Beyer et al. 2000). Thus, PET and CT images are acquired sequentially, reconstructed, and finally superimposed to produce combined images that provide accurate anatomical information regarding abnormal lesions detected by PET. This is inherently limited in the stand-alone PET because it has low spatial resolution and signal-to-noise ratio, and thus insufficient anatomical information. In addition to providing anatomical information, CT also provides information on photon attenuation, and the resulting CT-based attenuation correction reduces whole-body scan times by up to 40% and provides essentially noiseless attenuation correction factors (Townsend 2008).

Shortly after the introduction of the first commercial integrated PET/CT systems in 2000, they overtook stand-alone PET systems in terms of numbers sold.

### 7.2.2 MRI

MRI is a truly amazing imaging technique that continues to defy the limitations placed on the spatial resolutions of conventional optical instruments by diffraction. If its spatial resolution were limited by diffraction, it would achieve resolution on the order of several meters, that is, at approximately the wavelength of the radiofrequency (RF) radiation used. MRI achieves its exquisite spatial resolution by locating the origins of RF waves to much smaller volumes than those allowed by RF wavelengths by using nuclear magnetic resonance.

Nuclei with a spin of ½, such as, $^{1}H$, $^{13}C$, and $^{19}F$, can adopt two energy states in a magnetic field, namely, a low-energy "spin-up" state and a high-energy "spin-down" state; the populations of nuclei in these states can be determined using the Boltzmann relation. Furthermore, nuclei in the low-energy state are "flipped" to the higher-energy state by RF of energy equal to the difference between the energies of the spin-up and spin-down states, and nuclei that are flipped into the spin-down state subsequently return to the spin-up state by emitting an RF photon at a time determined by the surrounding material. Thus, MRI can better differentiate soft tissues than CT. Furthermore, in addition to the magnetic field and RF waves, the origins of RF waves are also located using a magnetic field gradient, which restricts the region of magnetic resonance. Thus, depending on how and when RF waves and gradient

fields are applied (commonly referred to as *pulse sequences*), MRI can produce quite different images of the same subject.

The excellent spatial resolution of MRI and its ability to differentiate soft tissues and surrounding material have resulted in its preferred use for applications as diverse as disease diagnosis, cognitive brain functional investigations, and molecular imaging. Nevertheless, the sensitivity of MRI is five or six orders lower than that of PET, because differences between the fractions of nuclei in high- and lower-energy states are on the order of one per million, even at magnetic field strengths of several tesla. Research activities in the MRI area are focused on increasing sensitivity using high magnetic fields, reducing scan times, and broadening its field of application to molecular imaging by using nanoparticle-based contrast agents.

### 7.2.3 PET/MRI

Although many of the benefits of PET/CT over stand-alone PET are well appreciated, several limitations of PET/CT have also been identified. Physiological and voluntary motions of the body and organs that occur between CT and PET scans can cause artifacts in fused images due to the sequential scan protocols used; simultaneous acquisition is not allowed because the high flux of X-ray photons required for CT imaging can saturate PET scintillation detectors.

Higher levels of radiation exposure during PET/CT examinations as compared with stand-alone PET or CT are also problematic, although the currently performed PET/CT studies are justified on the basis of potential medical benefit. The additional radiation dose requirement imposed by CT can be reduced to under 5 mSv if low-dose CT is performed for attenuation correction and anatomical localization only. Nevertheless, this amount of radiation is still of concern; for example, the current ICRP (International Commission on Radiological Protection) recommended limit for the general public is 1 mSv/year.

The soft tissue contrast provided by CT is considerably lower than that provided by MRI, and this higher contrast provided by MRI means better diagnostic accuracies, particularly in brain studies. Furthermore, it is useful for local tumor assessments and whole-body staging, where the higher soft tissue contrast of MRI is beneficial.

In principle, simultaneous PET and MR imaging is possible, because MRI utilizes nonionizing RF waves, which do not affect PET scintillation crystals, and conversely, PET gamma rays are not detected by MRI receiver coils. Thus, simultaneous PET/MR imaging has several obvious advantages over PET/CT. In particular, correlative assessments of function and anatomy, and multidimensional evaluations of functional, biochemical, and molecular processes are likely to be possible using PET/MRI. The simultaneous assessment of metabolism or neuroreceptor/ligand interactions by PET and biochemical contents by MR spectroscopy (MRS) is an example, and comparative studies of cerebral blood flows determined by $H_2{}^{15}O$ PET and BOLD (blood-oxygen-level dependent) signals determined by functional MRI (fMRI) will undoubtedly provide valuable information on the physiological basis of cognition in the brain. Furthermore, dual-modality molecular imaging with several imaging probes introduces many other possibilities.

Furthermore, because it requires less radiation exposure than PET/CT, PET/MR imaging would be particularly useful for pediatric studies and for treatment monitoring. Breast-specific PET/MRI is also a better proposition than PET/CT for diagnostic and screening purposes, because breast tissue is highly sensitive to penetrating radiation.

## 7.3 OTHER PHOTODETECTORS USED IN PET/MRI UNITS

### 7.3.1 The Photomultiplier Tube

#### 7.3.1.1 Structure and Principles

From a technical perspective, scintillation detectors, which consist of inorganic scintillation crystals and photodetectors, are of central importance to PET. Visible and ultraviolet (UV) photons emitted by these crystals should be converted into electrical energy by low-noise, high-internal-gain photodetectors, which are necessary to reduce the complexity and cost of the electronic readout circuits. Currently, photomultiplier tubes (PMTs) are usually used as photoelectric converters in PET scanners. These units convert light energy into electron movement using a photocathode, and this electrical current is subsequently amplified by applying high voltages between a series of cascading dynodes.

The main advantage of PMTs is the high gain achieved ($10^6$–$10^7$), which enables the use of standard electronics for signal processing. PMTs also have excellent timing properties; the rise time of signal output range from 0.5 to 3 ns, and variances in electron transit times between the photocathode and anode can be reduced to less than 300 ps. PMTs are also relatively insensitive to temperature and operating voltage changes.

However, PMTs are expensive to manufacture, because mechanical components within PMT vacuum tubes are largely handmade. Furthermore, PMTs are highly sensitive to magnetic fields, which is of obvious concern when one is considering PET/MRI or combination techniques involving scintillation cameras and MRI. Essentially, Lorentz forces, generated by applied magnetic fields, deflect photoelectrons from their original trajectories and much reduce output signal strengths. The suitability of PMTs for use in PET/MRI applications is also limited by the space available within MRI scanners. For these reasons, the placement of PET detectors within the MR magnet is not a practical possibility, though several methods of reducing magnetic-field signal disruption and PET data-acquisition interference have been investigated.

#### 7.3.1.2 Integration of Separate Scanners

A PET/MRI system with the current PET/CT configuration (sequential scans using a parallel machine arrangement) has been developed by Cho and colleagues at the Gachon Medical School, Korea (Cho et al. 2008). According to this approach, a Siemens HRRT PET (a brain-specific unit) and a 7.0-T MRI scanner are placed and operated independently in neighboring rooms. A mechanically precise table transfer system is used to shuttle patients between the units; the fronts of the two gantries face each other, and the table is rotated through 180° during transfer (Figure 7.2a). This design allows the two systems to be operated under optimal operating conditions without electromagnetic interference. However, this required the installation of

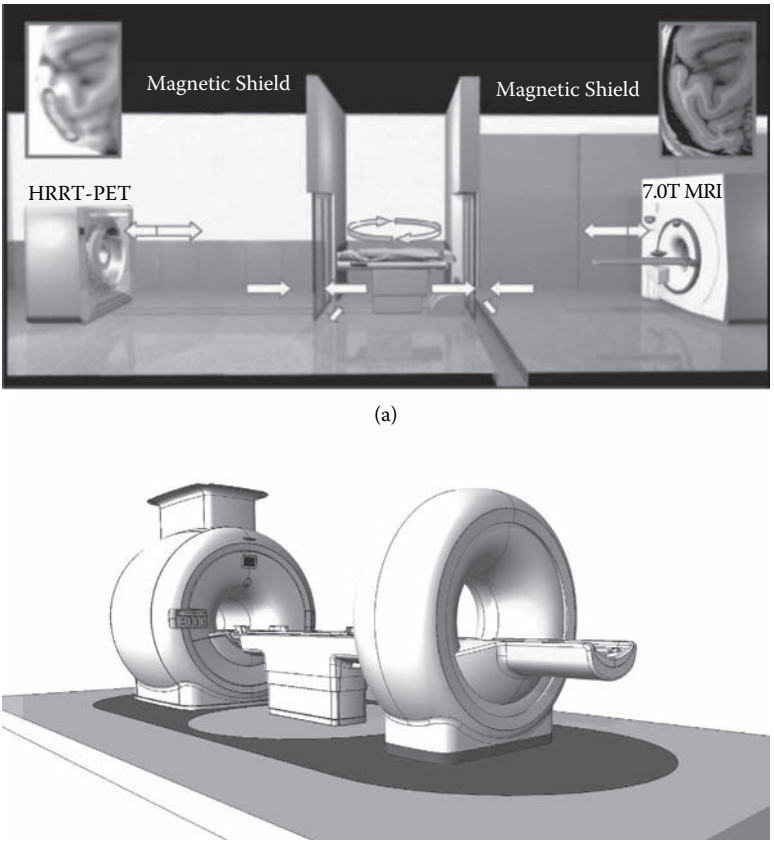



**FIGURE 7.2** The integration of PMT-based PET and a high-field MRI scanner; (a) brain-dedicated PET and 7.0T MRI scanners require neighboring rooms (Reprinted from Cho et al., 2008. With permission.), whereas (b) time-of-flight PET and a self-shielded 3.0T MRI can be accommodated in the same room. (© Philips Medical System. With permission.)

≈500 tons of steel magnetic shielding to reduce the strength of the field produced by the nonself-shielded superconducting magnet of the 7.0-T MRI unit to achieve under 0.5 gauss at the PET unit.

Philips Medical Systems adopted a similar design concept when it developed a clinical whole-body PET/MRI system. Time-of-flight (TOF) PET and 3.0-T MRI scanners, which are the most advanced systems among the clinically used equipment operated there, are combined (Schulz 2008) (Figure 7.2b). Because of the lower level of electromagnetic interference induced by the self-shielded 3.0-T MRI magnet and the use of additional shielding materials around individual PMTs and the PET gantry, these two scanners can be operated in one room, provided a minimum distance of ≈2.5 m is maintained between gantry surfaces. This system also uses a foot-to-foot arrangement and rotating bed, like the Gachon unit. Furthermore, because whole-

body imaging systems are integrated, this system can be applied in any clinical field, including oncology and cardiology.

A combination of whole-body PET/CT and 3.0-T MRI is also being undertaken at the Korea Institute of Radiological and Medical Sciences (KIRAMS). In contrast to the previously mentioned systems, the PET/CT and MRI scanners are located parallel to each other in adjoining rooms. This system is being developed to provide hybrid whole-body PET/MR or PET/CT/MR images clinically with high-accuracy image registration. The use of widely available whole-body PET/CT and MRI systems and minimal modification enables the lower cost development of hybrid whole-body imaging systems, as compared with the systems mentioned previously. This development at KIRAMS also involves the development of data acquisition and processing procedures required for fully integrated PET/MRI systems.

Currently, the shortcomings of these systems that receive the most attention are the longer acquisition time required and the lower patient throughputs associated with the use of separate scanners.

#### 7.3.1.3 Approaches Using Optical Fibers

The use of long optical fibers to isolate PMTs from strong magnetic fields offers another approach to the design of PMT-based PET systems. According to this scheme, only scintillation crystals are located inside the MRI magnet, and scintillation light is transferred via optical fiber to PMTs located outside (Figure 7.3).

Shao et al. (1997) proposed this method and acquired PET/MR images with a prototype detector system using 4-m-long double-clad optical fibers connected to the sides of $2 \times 2 \times 10$ mm$^3$ LSO crystals in a crystal ring and the individual pixel of a multichannel PMT; PMTs and readout electronics were enclosed in a shielded box. Similarly, Marsden et al. (2002) improved PET detection sensitivity without degrading resolution uniformity by arranging LSO crystals, coupled to PMTs with optical fibers, in four concentric rings (Figure 7.3a).

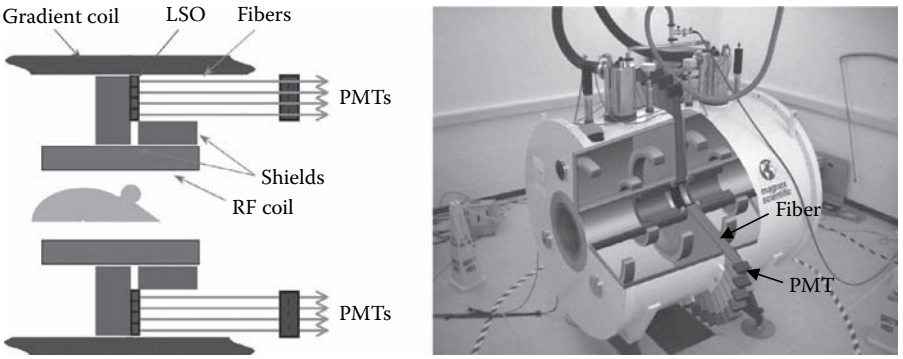



**FIGURE 7.3** Examples of PET/MRI designs using optical fibers between scintillation crystals and PMTs: (a) light readout from the side of a crystal, (b) readout from the back surface of a crystal. (Reprinted from Marsden et al., 2002 and Lucas et al., 2006. With permission.)

Using these approaches, only one imaging slice was possible because optical fibers were attached to the side surfaces of the crystals. To overcome this limitation, Yamamoto et al. (2005) proposed the use of a block detector that allowed the axial and radial interaction positions of scintillation crystals (slice location and depth-of-interaction [DOI] positions, respectively) to be measured for multislice imaging. In this approach, the light-sharing method devised for jPET at the National Institute of Radiological Science in Japan was applied (Orita et al. 2005). This method allows slice numbers to be further increased by combining the light-sharing method with other DOI measurement methods, such as that involving the stacking of multiple scintillation crystals with different decay times (Inadama et al. 2006; Hong et al. 2008a). However, the axial extent of this block detector is limited to a few centimeters because of increased light output nonuniformity due to the greater total lengths of the crystals used.

Researchers at West Virginia University and at the Thomas Jefferson National Accelerator Facility coupled 90 degree bending, 2.5-m-long optical-fiber bundles to the backs of crystals of dimension $5 \times 5 \times 1.5$ cm$^3$ in order to increase the number of slices and the axial field of view (Raylman et al. 2006). They used a flat-panel PMT less sensitive to magnetic fields due to its compact dynode structure, and also added a proximity electron-focusing scheme.

The advantage of the optical fiber approach is that it allows the use of proven, robust PMT-based scintillation detectors with minimal modification. However, it degrades energy, timing, and spatial resolutions due to light losses during optical transmission, which can reduce PET image quality because these performance reductions increase the rates of scattered, random, and mispositioned coincidences.

Thus, the minimization of light losses caused by the lengths of and bends in optical fibers is the most important outstanding technical issue in the optical fiber approach. One of the methods used to reduce fiber length involved the use of a split magnet (Lucas et al. 2006) (Figure 7.3b). Within the 80-mm gap between the actively shielded 1.0-T magnet structure, PET detector modules of a Siemens microPET Focus 120 system (Kim et al. 2007) were inserted after design modification. The original 10-cm-long optical bundle was replaced with a 120-cm bundle, and PMTs were then located outside the magnet, where the magnetic field strength was ≈30 mT. The use of this relatively shorter optical-fiber length, as compared to previous approaches, allowed better energy and timing resolutions (25% and 3.6 ns, respectively) to be obtained (Hawkes et al. 2008).

### 7.3.2 Avalanche Photodiode

#### 7.3.2.1 Structure and Principles

Semiconductor photodetectors have several advantages over PMTs for PET/MRI use because they are not sensitive to magnetic fields, are compact, and have high quantum efficiencies. PIN (p-i-n) diodes are rarely used in PET scintillation detectors because they do not have internal gain, whereas avalanche photodiodes (APDs) are similar to a typical PIN diode but have a more complicated internal structure, and they amplify scintillation signals by utilizing a cascade of electron-hole pairs that are created by photons impacting the p-layer of APDs. The free electrons generated

are accelerated by a high potential (100–200 V in silicon) and generate secondary electron-hole pairs, which generate an electron avalanche. Nevertheless, APDs only offer a gain of $\approx 10^3$, which is much smaller than $\approx 10^6$, the gain of PMTs, and thus APDs require low-noise amplifiers.

#### 7.3.2.2 Array-Type APDs

Several groups, including the Technische Universität München, Sherbrooke University, and the Brookhaven National Laboratory, have devised small-animal PET imaging systems based on array-type APDs. APD arrays are suitable for small-animal PET systems that require fine spatial resolution because APDs can be coupled to individual crystals. LabPET was the first APD-based PET scanner commercialized and was developed by the Sherbrooke University group and commercialized by Gamma Medica and General Electric.

The APD detector has also attracted interest as a means of developing PET/MRI systems, because it is compact and unaffected by magnetic fields (Figure 7.4). The sizes of PET systems used in PET/MRI units are important because magnet prices are highly dependent on magnet bore.

Pichler et al. (1997) showed that the LSO-APD detector is suitable for PET/MRI applications if bias voltage and temperature are stabilized, and they found that the gain and energy resolutions of their LSO-APD detector were unaffected by a 9.4-T magnetic field. In addition, the direction of the electrical field of p–n crossing of the APD relative to the magnetic field had no effect. An MR-compatible PET insert based on the LSO-APD detector was developed by this group for preclinical imaging using a 7.0-T MRI scanner (Pichler et al. 2006; Judenhofer et al. 2008). PET detector modules—consisting of a $12 \times 12$ crystal array, a $3 \times 3$ APD array, and preamplifier electronics—are enclosed in individual copper-shielding housings and inserted between the gradient and RF coils. Simultaneous PET/MR imaging was possible

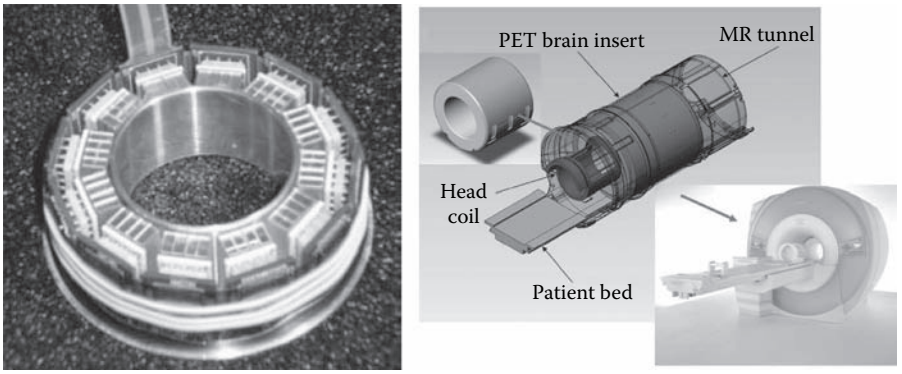



**FIGURE 7.4** APD-based PET inserts for simultaneous PET/MR imaging; (a) for small animal imaging and (b) for human brain imaging. (Reprinted from Woody et al., 2007 and Schlemmer et al., 2008. With permission.)

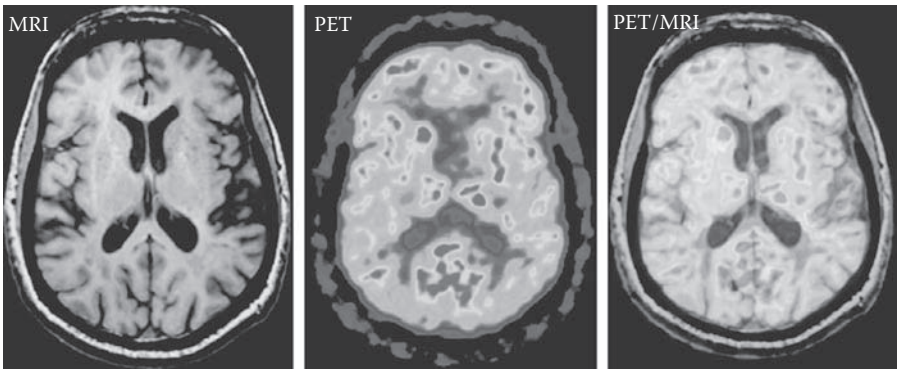



**FIGURE 7.5** Human brain images acquired simultaneously using APD-based PET inserts. (Reprinted from Schlemmer et al., 2008. With permission.)

using this system without significantly degrading the physical performance or image quality of either imaging modality.

Researchers at the Brookhaven National Laboratory also developed an APD PET insert based on the technology used for their RatCAP (Rat Conscious Animal PET) scanner (Woody et al. 2007) (Figure 7.4a). This system was originally designed for the simultaneous acquisition of the PET and MR images of small-animal brains, and is currently being extended to breast imaging applications (Ravindranath et al. 2008).

Siemens Medical Systems has also integrated APD detector technology with a clinical 3.0-T MRI unit to enable the simultaneous and isocentric PET/MR imaging of the human brain (Schlemmer et al. 2008) (Figure 7.4b, Figure 7.5), and is currently developing a whole-body PET/MRI system.

#### 7.3.2.3 Position-Sensitive APDs

Position-sensitive APDs (PSAPDs) provide intrinsic position-sensing capability, which enables the number of output channels to be significantly reduced without additional signal multiplexing. The PSAPDs manufactured by Radiation Monitoring Devices, Inc., collect signals from four contacts placed at the corners of the resistive layer of a high-gain planar APD to compute event positions in coupled scintillation crystals (Shah et al. 2002).

This PSAPD has been used in an MR-compatible PET insert developed by the University of California at Davis group for small-animal imaging (Catana et al. 2006). Two major technologies (optical fiber readout and semiconductor detectors), which are being investigated for simultaneous PET/MR imaging, were also incorporated into this development. The LSO crystal array used was coupled using short, bent optical fiber bundles to a PSPAD, which was placed outside the RF coil and copper shielded with other electronics. The feasibility of the LSO-PSAPD detector for small-animal imaging was demonstrated using a 7.0-T magnet. Although pincushion distortion was encountered due to the four-corner readout system, crystal identification was possible in the flood histogram, and energy resolution was acceptable. However, significant variations in timing resolution across the PSAPD surface required the use of a wide

coincidence-timing window (40 ns). Improvements in these shortcomings are required before this detector module is suitable for human imaging studies.

## 7.4 GEIGER-MODE APD

### 7.4.1 Structure and Principles

Since the original Russian development of Geiger-mode APDs (G-APDs) in the 1980s, several institutions and commercial organizations have been working on different designs (Golovin and Saveliev 2004; Saveliev 2004; Yamamoto et al. 2007; Moser et al. 2007). G-APD microcells basically consist of a p–n junction, which is used to generate electron-hole pairs, and a quenching resistor to control the avalanche process. Different G-APD designs differ with respect to p–n junction disposition and the control of avalanche quenching. Figure 7.6 illustrates a cross-sectional view of a G-APD microcell in which the Geiger junction is isolated from the p-substrate (Stapels et al. 2006). This substrate prevents breakdown at the edge of the 30-μm-diameter photodiode pixel used when it is operated in Geiger mode. If the self-propagating avalanche is unchecked, it causes the diode to conduct too much current. On the other hand, if a circuit element detects the presence of an avalanche current and drops the bias below the critical bias voltage, the self-propagating avalanche is quenched. Subsequently, the bias can be increased to above the threshold to await the arrival of another single-photoelectron event, thereby resetting the G-APD pixel (Stapels et al. 2006).

G-APDs have 500–4000 microcells/mm$^2$ connected in parallel, which produce one common output signal, and although individual Geiger-mode microcells do not provide proportional output when combined in this way, they do provide a signal that is proportional to the energy of incident radiation.

### 7.4.2 Basic Characteristics of G-APDs

Because of the different requirements of the various G-APD applications—for example, as calorimeters in high-energy physics experiments and as Cherenkov detectors in astrophysics—we present the characteristics required for G-APDs for PET applications only. Information on other important characteristics, such as radia-

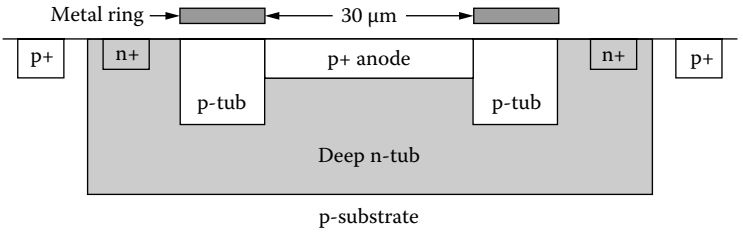



**FIGURE 7.6** A cross-sectional view of G-APD. (Reprinted from Stapels et al., 2006. With permission.)

tion hardness, optical cross talk, and dark counts, can be found in an excellent review by Renker and Lorenz (2009).

### 7.4.2.1 Bias Voltage and Gain

Because G-APDs operate in the Geiger region and use bias voltages above the breakdown voltage to achieve high gain, gain sensitively depends on bias voltage, and thus even a small dependency of bias voltage on temperature can be problematic. For example, the gains of the Hamamatsu G-APD (MPPC: multipixel photon counter) exponentially vary from ≈10%/0.1 V for the S10362-11-050U/C and ≈30%/0.1 V for the S10362-11-100U/C (Figure 7.7). Typical gains are $5 \times 10^5$ at 70.1 V for the S10362-11-050U/C and $2 \times 10^6$ at 70.1 V for the S10362-11-100U/C, and bias voltage changes by about 1 V for both when the temperature is increased from 20°C to 30°C (Yamamoto et al. 2007). Similar dependencies of gain on bias voltage and on temperature have been reported for other G-APD devices (Musienko, Reucroft, and Swain 2007; Ramilli 2008). The combined effect of sensitivity and temperature dependency could seriously destabilize the G-APD PET system. Furthermore, the gain change due to a fractional bias change is about 10 times larger for a G-APD than for a PMT, based on a rate of 50%/50 V at about 950 V for the Hamamatsu H9500 flat-panel PMT.

In addition, the optimal bias voltages of G-APDs vary considerably. To achieve high spatial resolutions and optimize manufacturing cost, most PET systems use Anger type or charge division electronics that allow the use of smaller crystals than G-APDs. In this configuration, events from a number of crystals are read by multiple G-APDs, and crystals struck by gamma photons are identified using the output signals of G-APDs. Therefore, minimal variation in G-APD output is required to identify crystals clearly and to allow the rejection of random events and noise.

As mentioned previously, sensitivity to bias voltage and variations in optimal bias voltage are likely to limit the usefulness of the G-APD PET system, unless the

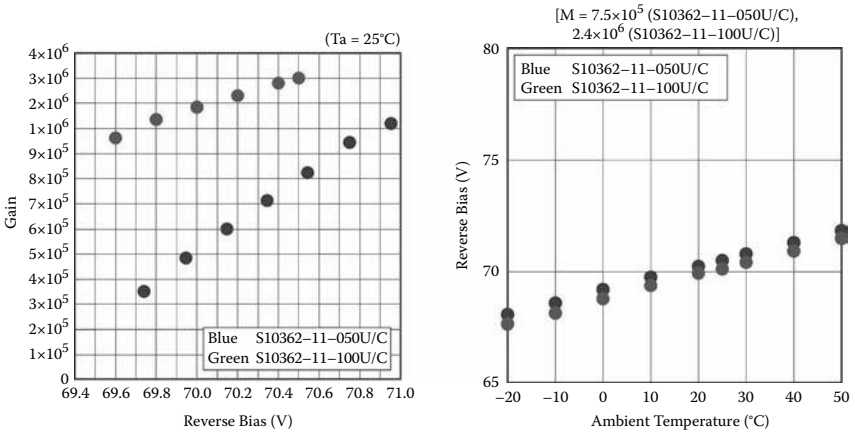



**FIGURE 7.7** Gain dependency of a G-APD with reverse bias as a function of ambient temperature. (© Hamamatsu Photonics. With permission.)

bias voltage to each G-APD can be controlled precisely and dynamically (due to the dependence of optimal bias on temperature).

The latest MPPC S10985-050C contains $2 \times 2$ sensitive cells of $3.0 \times 3.0$ mm$^2$/cell. However, our tests show that four cells cannot be biased using a single voltage source because of large variations in the photopeaks of 511-keV gamma rays.

#### 7.4.2.2 Energy and Timing Resolutions

Since G-APDs consist of a few hundred to a few thousand microcells that all operate in the Geiger region, measured energy is proportional to the number of scintillation events. If the number of events is excessive, several photons are likely to enter individual microcells, which would cause a loss of proportionality between the output signal and radiation energy. Therefore, manufacturers recommend that G-APDs be operated such that fewer than 60% of microcells fire simultaneously. However, because most of the G-APDs used for PET have many microcells, these G-APDs are unlikely to be saturated by 511-keV photons. It is also worth noting that G-APDs and PMTs coupled to same scintillation crystals have similar energy resolutions of 10%–20%.

One of the advantages of G-APD over APD is that its response time is shorter (<1 ns versus a few nanoseconds), which results in excellent time resolution when coupled with fast crystals. In fact, several groups have reported subnanosecond resolutions (Yamamoto et al. 2007; Schaart et al. 2009). However, the large capacitances of MPPCs result in slow pulses that require a low trigger threshold to achieve good time resolution (Kim, Wang, and Dolinsky 2008; Nassalski et al. 2008).

### 7.4.3 MR Compatibility

G-APDs are intrinsically as compatible with MRI units as APDs are. The p–n junction of a G-APD where an avalanche occurs is very narrow (less than 10 μm), and the electrons and holes produced by events are subjected to electric fields of >100,000 V/cm and move only short distances between the two electrodes. On the other hand, taking the Hamamatsu flat-panel H9500 PMT as a typical example, electron transition lengths are on the order of a few centimeters, and electrons in PMTs are subject to electric fields of only ≈1000 V/cm. For this reason, the electrons and holes generated within a G-APD are unaffected by the magnetic and RF fields generated by MR units.

## 7.5 MR-COMPATIBLE G-APD PET DETECTORS

### 7.5.1 Approaches Used at Seoul National University

A team from Seoul National University (SNU), composed of biomedical engineers and nuclear medicine and MR physicists, has studied MRSs (metal resistor semiconductors), SSPMs (solid-state photomultipliers), and MPPCs since 2006. Because the quantum efficiency of MRSs produced in 2006 was poor for blue light, signal amplitudes were low and energy resolutions were poor. SSPMs with much higher quantum efficiencies and blue sensitivities were studied during early 2007 (Lee et al. 2007) and, due to their excellent energy and time resolutions, were subsequently

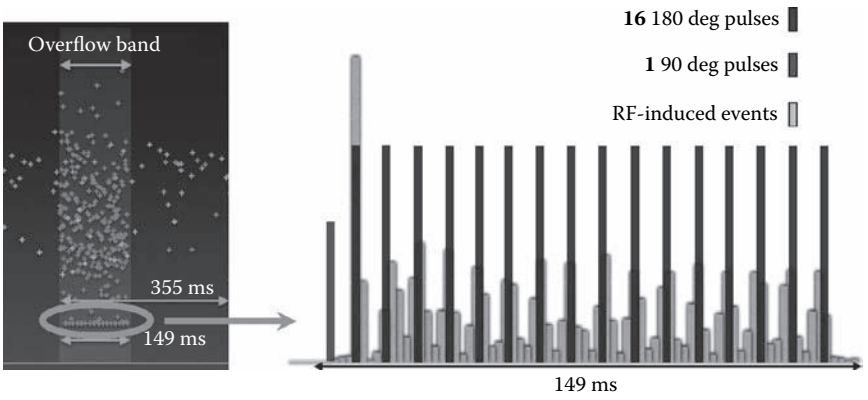



**FIGURE 7.8** RF-induced events matched with pulse sequences during a T2 turbo pulse sequence.

investigated in terms of MR compatibility. Two SSPMs coupled with LYSO crystals irradiated with a $^{22}$Na radiation source were positioned inside a 3.0-T MRI, and the coincident signals from two SSPMs were read out using VME and NIM modules while the MRI unit was scanning a phantom using gradient and spin-echo pulse sequences. The excellent energy and coincident time resolutions obtained were found to be unaffected by these MRI sequences, which verified that SSPMs are unaffected by RF pulses (Hong et al. 2008b).

Conductive materials block magnetic and RF fields, but on the other hand, RF waves induce electric currents inside conductive materials, and these currents produce noise in PET images. Shielding materials of thickness ≈0.1 mm induced huge noises during RF sequences and increased event counts substantially despite the setting of trigger thresholds at levels high enough to capture only the photoelectric events due to 511-keV photons. Figure 7.8 shows that RF-induced events in PET electronics match one 90 degree RF pulse and sixteen 180 degree RF pulses during T2 turbo RF sequences (Lee et al. 2008). Even though it may be possible to exclude these unwanted events by overriding the false trigger during RF application, this would be undesirable because of the dead-time loss caused in the front-end electronics of PET detectors. Thus, because induced currents in shielding materials are proportional to shield areas and thicknesses, a modularized PET system that requires only thin shielding material for RF protection is preferable.

As mentioned previously, because of sensitivity to bias voltage and temperature, bias voltages supplied to G-APDs must be precisely and individually controlled when G-APD systems use Anger type or charge division electronics to achieve spatial resolutions that exceed G-APD pixel dimensions (Figure 7.9). The SSPM block shown in Figure 7.9 consists of 4 SSPMs, 16 LGSO $1.5 \times 1.5 \times 7.0$ mm$^3$ crystals, and a quartz crystal layer of thickness 3.0 mm. Each SSPM has a packaging area of $4.0 \times 4.0$ mm$^2$ and a sensor area of $2.1 \times 2.1$ mm$^2$. At SNU, we have tested more than 100 SSPMs and encountered high sensitivities to bias voltage and large optimal bias-voltage variations. A digitally controlled bias-voltage supply system that also

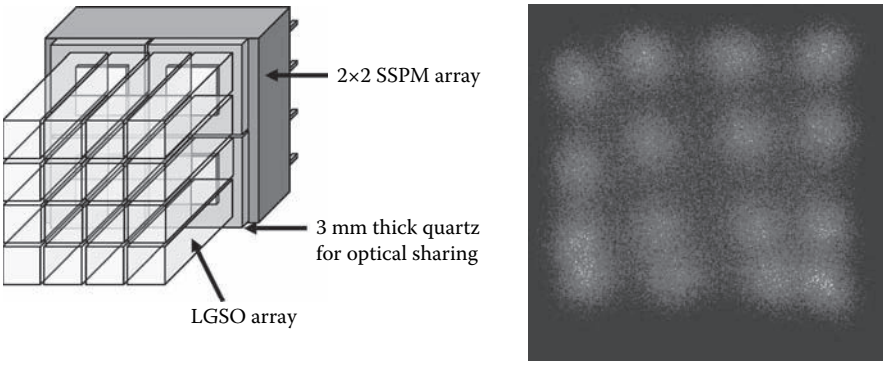



**FIGURE 7.9** SSPM and crystal array to obtain a finer spatial resolution than SSPM dimension. Each SSPM has a packaging area of $4.0 \times 4.0$ mm$^2$ and a sensor area of $2.1 \times 2.1$ mm$^2$.

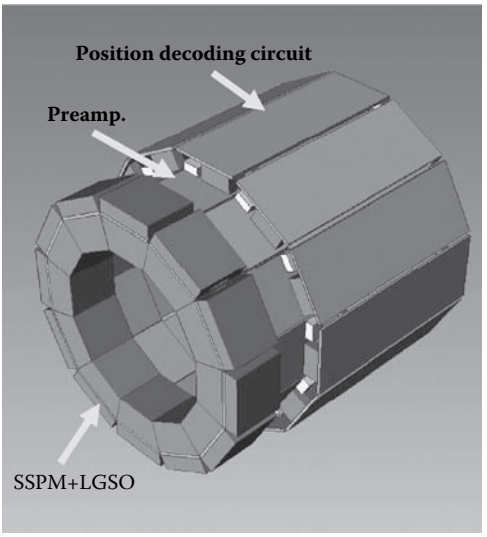



**FIGURE 7.10** A schematic of a modularized G-APD PET system being developed at Seoul National University.

monitors temperature is currently being developed for the precise and individual control of bias voltages (Kwon et al. 2009).

Figure 7.10 illustrates a G-APD PET system being developed at SNU with a modular structure that was adopted to reduce RF-induced currents and to allow digital control of bias-voltage supply.

### 7.5.2 TOF PET/MRI

The fast response time of G-APDs is an attractive characteristic, and this has been adopted as a research topic for the development of advanced PET devices (Figure 7.11). Time-of-flight (TOF), defined as the difference between the arrival

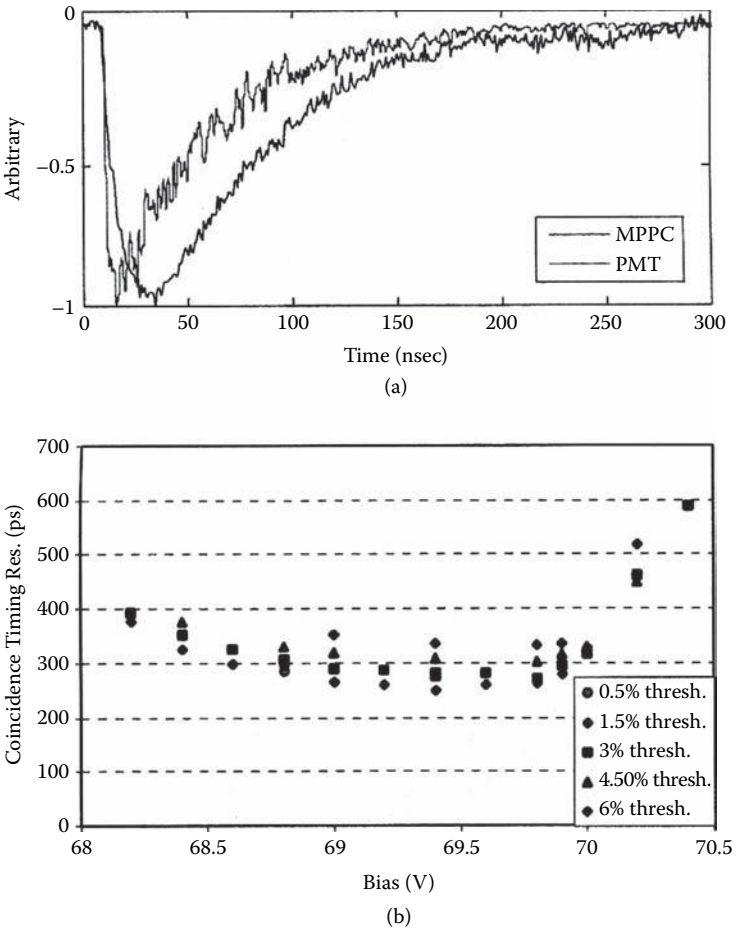



**FIGURE 7.11** Potential of G-APD for time-of-flight PET; (a) comparison of pulse shapes from an MPPC and a PMT, and (b) coincidence timing resolutions measured using LYSO/MPPC pairs are shown as functions of bias voltage and trigger level. (Reprinted from Kim et al., 2008.With permission.)

times of two gamma rays at a detector, is useful for reducing noise far from the annihilation point in reconstructed images. Furthermore, due to the lower noise levels of TOF PET images, radio-pharmaceutical injection doses and scan times can be reduced. The Philips Gemini-TruFlight unit is the most popular commercial TOF PET based on fast PMT and has a timing resolution of ≈600 ps (Surti et al. 2007). The Philips company is also leading a European consortium (http://www.hybrid-pet-mr.eu) to develop a whole-body PET/MRI system with TOF capability by utilizing G-APDs as fast photodetectors. The SiPM-based G-APD from FBK-irst (Trento, Italy)—optimized for fast response, photon detection efficiency (PDE), and packaging—is also being used in this development.

A highly integrated low-power-readout ASIC (application-specific integrated circuit) is also being investigated to analyze signals from individual SiPMs. The most recent version is a 40-channel time and energy readout ASIC designed in a 0.18-μm CMOS (complementary metal oxide semiconductor) process; this ASIC yielded intrinsic coincidence-timing resolutions of 54-ps FWHM (full width at half maximum) for two channels on chip and 69-ps FWHM between chips (Ritzert et al. 2008). Using this ASIC, an energy resolution of 16.1% for 511-keV photons was measured using LYSO crystals and a $3 \times 3$ mm$^2$ FBK-irst SiPM.

Several other groups are also investigating the feasibility of G-APD for TOF measurements and have reported timing resolutions similar to or better than PMT-based measurements (Kim, Wang, and Dolinsky 2008; Vinke et al. 2009; Llosa et al. 2007). However, because coincidence-timing resolution is dependent on bias voltage and temperature, control of these parameters is critical to achieve adequate G-APD timing properties. In addition, the slower increases of G-APD output pulses versus PMT (due to the slow recovery of microcells caused by passive quenching resistance) require low trigger levels for timing pickup to obtain optimal timing resolution, and these low trigger levels make the detector sensitive to noise and dark counts (Kim, Wang, and Dolinsky 2008).

### 7.5.3 DOI PET/MRI

Fine spatial resolution is essential for small-animal imaging during preclinical research. This improved spatial resolution is also valuable during clinical studies because lesion detectability is enhanced. PET spatial resolution can be improved by reducing the data sampling distance with a scintillation crystal element that has a smaller surface area and by minimizing the effect of noncollinear gamma ray annihilation by reducing the crystal ring diameter as much as possible. On the other hand, the length of crystal elements should not be much reduced to maintain the detection efficiency for gamma rays.

The cost of the spatial resolution and sensitivity improvements achieved by adopting such configurations is the rapid degradation of spatial resolution in the peripheral region of the field of view. In this region, gamma rays that obliquely enter the detector ring can pass over several detector elements without interacting, and are finally detected by elements that do not correspond with the original incidence position. This positioning error (parallax error) is of importance during PET/MRI development because PET inserts integrated into an existing MRI system for imaging small animals or peripheral organs have small ring diameters. In fact, they are usually smaller than the previous stand-alone PET scanners used for the same purposes because of space limitations imposed by the MR magnet and coils on the size of the PET detector ring.

Several suggestions have been made concerning the measurement of depth of interaction (DOI) in crystals as a means of correcting PET parallax error, and these are summarized in a review by Lewellen (2008). The compact size of semiconductor detectors is useful for acquiring DOI information in some schemes. A DOI detector design in which multiple layers of a monolithic crystal (a continuous crystal slab) and a G-APD array are stacked to give intrinsic DOI information has been suggested

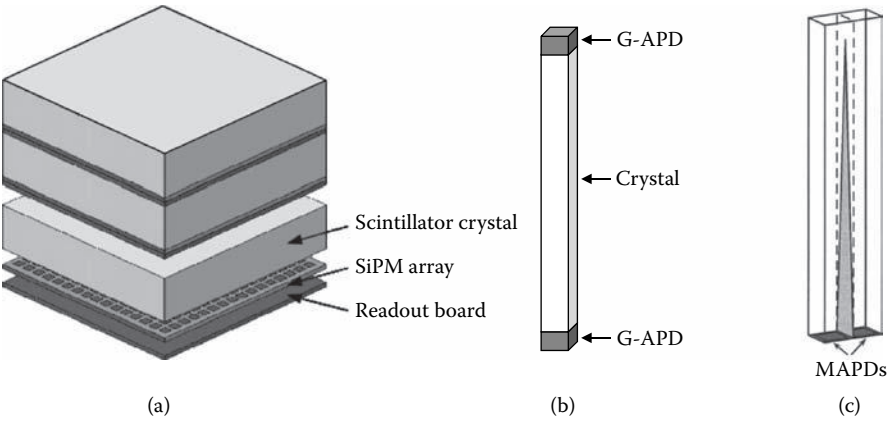



**FIGURE 7.12** Schemes of depth-of-interaction (DOI) measurements using a G-APD: (a) stacking multiple layers of scintillation crystal and G-APD array coupling, (b) dual-ended readout scheme, and (c) single-ended readout scheme with light spread tailoring. (Reprinted from Herbert et al., 2007 and Champley et al., 2008 .With permission.)

(Herbert et al. 2007) (Figure 7.12a). Furthermore, it has also been shown that DOI encoding by measuring light output from both ends of a crystal using a G-APD pair and calculating the ratios of these signals is possible (Shao, Li, and Gao 2007) (Figure 7.12b). DOI resolution measured using this method and a $1.8 \times 2 \times 20$ mm$^3$ LSO crystal and SSPM pair (sensitive area = $1 \times 1$ mm$^2$, peak wavelength = 40 nm, PDE = 20%) was 4.5 mm. The University of Washington group used a different light-sharing scheme with a G-APD pair (Lewellen et al. 2007). According to this system, light sharing between two adjacent crystal elements is tailored using a triangular light reflector, as shown in Figure 7.12c, and DOI dependence on the ratio of light outputs from crystal ends is used to estimate the DOI position. The obvious advantage of this method over the dual-ended readout scheme is that the number of photodetectors and readout channels can be halved.

## 7.6 CONCLUSION

The development of combined PET/CT has been achieved more by determination than technology, because both PET and CT technologies were mature, and no new technological breakthrough was required to combine the two. However, combined PET/MRI, especially with simultaneous imaging capability, appears to have followed quite a different developmental path. In particular, an MRI-compatible PET system has proven to be a difficult technical target, because the PMTs of PET units are extremely vulnerable to the magnetic fields generated by MRI systems. Moreover, the optical-fiber-based systems used to transfer scintillation light to a PMT positioned outside the magnetic field are costly in terms of the PET energy and time resolutions. However, solid-state devices, which are insensitive to magnetic and radiofrequency fields, have been developed, and these have presented new possibilities of combining PET and MRI. Furthermore, highly successful combined PET/MRI units with

excellent energy resolution have been constructed using APD units. In particular, the most recent development, G-APD, may allow the construction of combined PET/MRI with excellent timing and energy resolutions with TOF capability. However, in view of the fact that solid-state APD and G-APD devices remain inferior to PMT units, particularly in terms of stability and cost, combined PMT-PET/MRI based on some form of innovative arrangement of essentially separate units is likely to play an important practical role until a truly economic, solid-state PET/MRI is devised.

## REFERENCES

Beyer, T., D. W. Townsend, T. Brun, P. E. Kinahan, M. Charron, R. Roddy, J. Jerin, J. Young, L. Byars, and R. Nutt. 2000. A combined PET/CT scanner for clinical oncology. *J. Nucl. Med.* 41:1369–1379.

Catana, C., Y. Wu, M. S. Judenhofer, J. Qi, B. J. Pichler, and S. R. Cherry. 2006. Simultaneous acquisition of multislice PET and MR images: Initial results with a MR-compatible PET scanner. *J. Nucl. Med.* 47:1968–1976.

Champley, K. M., T. K. Lewellen, L. R. MacDonald, R. S. Miyaoka, and P. E. Kinahan. 2008. Statistical three-dimensional positioning algorithm for high-resolution dMiCE PET detector. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 4751–4754.

Cho, Z. H., Y. D. Son, H. K. Kim, K. N. Kim, S. H. Oh, J. Y. Han, I. K. Hong, and Y. B. Kim. 2008. A fusion PET-MRI system with a high-resolution research tomograph-PET and ultra-high field 7.0 T-MRI for the molecular-genetic imaging of the brain. *Proteomics* 8:1302–1323.

Golovin, V., and V. Saveliev. 2004. Novel type of avalanche photodetector with Geiger mode operation. *Nucl. Instr. Meth. A* 518:560–564.

Hawkes, R. C., T. D. Fryer, A. J. Lucas, S. B. Siegel, R. E. Ansorge, J. C. Clark, and T. A. Carpenter. 2008. Combined microPET®/MR system: Performance assessment of the full PET ring with split gradients. Paper presented at MR-PET Workshop, Jülich Germany.

Herbert, D. J., S. Moehrs, N. D'Ascenzo, N. Belcari, A. Del Guerra, F. Morsani, and V. Saveliev. 2007. The silicon photomultiplier for application to high-resolution positron emission tomography. *Nucl. Instr. Meth. A* 573:84–87.

Hong, S. J., S. Il Kwon, M. Ito, G. S. Lee, K. S. Sim, K. S. Park, J. T. Rhee, and J. S. Lee. 2008a. Concept verification of three-layer DOI detectors for small animal PET. *IEEE Trans. Nucl. Sci.* 55:912–917.

Hong, S. J., I. C. Song, M. Ito, S. I. Kwon, G. S. Lee, K. S. Sim, K. S. Park, J. T. Rhee, and J. S. Lee. 2008b. An investigation into the use of Geiger-mode solid-state photomultipliers for simultaneous PET and MRI acquisition. *IEEE Trans. Nucl. Sci.* 55:882–888.

Inadama, N., H. Murayama, M. Hamamoto, T. Tsuda, Y. Ono, T. Yamaya, E. Yoshida, K. Shibuya, and F. Nishikido. 2006. Eight-Layer DOI encoding of 3-dimensional crystal array. *IEEE Trans. Nucl. Sci.* 53:2523–2528.

Judenhofer, M. S., H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Rocken, R. E. Nutt, H. J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler. 2008. Simultaneous PET-MRI: A new approach for functional and morphological imaging. *Nat. Med.* 14:459–465.

Kim, C. L., G.-C. Wang, and S. Dolinsky. 2008. Multi-pixel photon counters for TOF PET detector and its challenges. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 3586–3590.

Kim, J. S., J. S. Lee, K. C. Im, S. J. Kim, S. Y. Kim, D. S. Lee, and D. H. Moon. 2007. Performance measurement of the microPET focus 120 scanner. *J. Nucl. Med.* 48:1527–1535.

Kwon, S. I., S. J. Hong, M. Ito, H. S. Yoon, I. C. Song, and J. S. Lee. 2009. Development of SSPM-PET with digitally controlled bias voltage supply. *J. Nucl. Med.* 50, suppl. 2:355.

Lee, J. S., M. Ito, K. S. Sim, B. Hong, K. S. Lee, J. Muhammad, J. T. Rhee, G. S. Lee, K. S. Park, I. C. Song, and S. J. Hong. 2007. Investigation of solid-state photomultipliers for positron emission tomography scanners. *J. Korean Phys. Soc.* 50:1332–1339.

Lee, J. S., K. S. Park, D. S. Lee, C. W. Lee, J. K. Chung, and M. C. Lee. 2005. Development and applications of a software for functional image registration (FIRE). *Comput. Methods. Programs. Biomed*. 78:157–164.

Lee, J. S., I. C. Song, S. I. Kwon, M. Ito, H. S. Yoon, K. S. Sim, J. T. Rhee, G. S. Lee, K. S. Park, and S. J. Hong. 2008. Test of Geiger-mode SSPM PET modules for the simultaneous PET and MRI acquisition. Paper presented at IEEE Nucl. Sci. Symp. Med. Imaging Conf., Dresden, Germany.

Lewellen, T. K. 2008. Recent developments in PET detector technology. *Phys. Med. Biol.* 53:R287–R317.

Lewellen, T. K., L. R. MacDonald, R. S. Miyaoka, W. McDougald, and K. Champley. 2007. New directions for dMiCE: A depth-of-interaction detector design for PET scanners. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 3798–3802.

Llosa, G., N. Belcari, G. Collazuol, A. Del Guerra, S. Marcatili, S. Moehrs, C. Piemonte, T. A. Carpenter, R. C. Hawkes, A. J. Lucas, and J. W. Stevick. 2007. Solid state evolution. *Nucl. Eng. Int*. 52 (December):18–20.

Lucas, A. J., R. C. Hawkes, P. Guerra, R. E. Ansorge, R. E. Nutt, J. C. Clark, T. D. Fryer, and T. A. Carpenter. 2006. Development of a combined microPET®: MR system. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 2345–2348.

Marsden, P. K., D. Strul, S. F. Keevil, S. C. R. Williams, and D. Cash. 2002. Simultaneous PET and NMR. *Brit. J. Radiol.* 75:S53–S59.

Moser, H. G., S. Hass, C. Merck, J. Ninkovic, R. Richter, G. Valceanu, N. Otte, M. Teshima, R. Mirzoyan, P. Holl, and C. Koitsch. 2007. Development of back illuminated SiPM at the MPI semiconductor laboratory. Paper presented at Int. Workshop on New Photon Detectors, Kobe, Japan.

Musienko, Y., S. Reucroft, and J. Swain. 2007. Tests and performance of multi-pixel Geiger mode APDs. Paper presented at Int. Workshop on New Photon Detectors, Kobe, Japan.

Nassalski, A., M. Moszynski, A. Syntfeld-Kazuch, T. Szczesniak, L. Swiderski, D. Wolski, T. Batsch, and J. Baszak. 2008. Silicon photomultiplier as an alternative for APD in PET/MRI applications. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 1620–1625.

Orita, N., H. Murayama, H. Kawai, N. Inadama, and T. Tsuda. 2005. Three-dimensional array of scintillation crystals with proper reflector arrangement for a depth of interaction detector. *IEEE Trans. Nucl. Sci.* 52:8–14.

Pichler, B., E. Lorenz, R. Mirzoyan, W. Pimpl, F. Roder, M. Schwaiger, and S. I. Ziegler. 1997. Performance test of a LSO-APD PET module in a 9.4 Tesla magnet. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 1237–1239.

Pichler, B. J., M. S. Judenhofer, C. Catana, J. H. Walton, M. Kneilling, R. E. Nutt, S. B. Siegel, C. D. Claussen, and S. R. Cherry. 2006. Performance test of an LSO-APD detector in a 7-T MRI scanner for simultaneous PET/MRI. *J. Nucl. Med*. 47:639–647.

Ramilli, M. 2008. Characterization of SiPM: Temperature dependencies. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 2467–2470.

Ravindranath, B., S. H. Maramraju, S. S. Junnarkar, S. S. Southekal, S. P. Stoll, J. F. Pratte, M. L. Purschke, X. Hong, D. Bennett, K. Cheng, D. Tomasi, D. S. Smith, S. Krishnamoorthy, P. Vaska, C. L. Woody, and D. J. Schlyer. 2008. A simultaneous PET/MRI breast scanner based on the RatCAP. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 4650–4655.

Raylman, R. R., S. Majewski, S. K. Lemieux, S. S. Velan, B. Kross, V. Popov, M. F. Smith, A. G. Weisenberger, C. Zorn, and G. D. Marano. 2006. Simultaneous MRI and PET imaging of a rat brain. *Phys. Med. Biol*. 51:6371–6379.

Renker, D., and E. Lorenz. 2009. Advances in solid state photon detectors. *J. Instr*. 4:P04004.

Ritzert, M., P. Fischer, I. Peric, and T. Solf. 2008. A 40-channel readout ASIC for ToF-PET. *IEEE Nucl. Sci. Symp. Med. Imaging Conf. Record* 5041–5044.

Saveliev, V. 2004. The recent development and study of silicon photomultiplier. *Nucl. Instr. Meth. A* 535:528–532.

Schaart, D. R., H. T. van Dam, S. Seifert, R. Vinke, P. Dendooven, H. Lohner, and F. J. Beekman. 2009. A novel, SiPM-array-based, monolithic scintillator detector for PET. *Phys. Med. Biol*. 54:3501–3512.

Schlemmer, H. P. W., B. J. Pichler, M. Schmand, Z. Burbar, C. Michel, R. Ladebeck, K. Jattke, D. Townsend, C. Nahmias, P. K. Jacob, W. D. Heiss, and C. D. Claussen. 2008. Simultaneous MR/PET imaging of the human brain: Feasibility study. *Radiology* 248:1028–1035.

Schulz, V. 2008. Simultaneous time-of-flight PET/MR. Paper presented at MR-PET Workshop, Jülich, Germany.

Shah, K. S., R. Farrell, R. Grazioso, E. S. Harmon, and E. Karplus. 2002. Position-sensitive avalanche photodiodes for gamma-ray imaging. *IEEE Trans. Nucl. Sci*. 49:1687–1692.

Shao, Y., S. R. Cherry, K. Farahani, K. Meadors, S. Siegel, R. W. Silverman, and P. K. Marsden. 1997. Simultaneous PET and MR imaging. *Phys. Med. Biol.* 42:1965–1970.

Shao, Y. P., H. R. Li, and K. K. Gao. 2007. Initial experimental studies of using solid-state photomultiplier for PET applications. *Nucl. Instr. Meth. A* 580:944–950.

Stapels, C. J., W. G. Lawrence, F. L. Augustine, and J. F. Christian. 2006. Characterization of a CMOS Geiger photodiode pixel. *IEEE Trans. Electron Dev*. 53:631–635.

Surti, S., A. Kuhn, M. E. Werner, A. E. Perkins, J. Kolthammer, and J. S. Karp. 2007. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. *J. Nucl. Med*. 48:471–480.

Townsend, D. W. 2008. Dual-modality imaging: Combining anatomy and function. *J. Nucl. Med.* 49:938–955.

Vinke, R., H. Löhnera, D. R. Schaartb, H. T. van Damb, S. Seifertb, F. J. Beekmanb, and P. Dendoovena. 2009. Optimizing the timing resolution of SiPM sensors for use in TOF-PET detectors. *Nucl. Instr. Meth. A* 610:188–191.

Woody, C., D. Schlyer, P. Vaska, D. Tomasi, S. Solis-Najera, W. Rooney, J. F. Pratte, S. Junnarkar, S. Stoll, Z. Master, M. Purschke, S. J. Park, S. Southekal, A. Kriplani, S. Krishnamoorthy, S. Maramraju, P. O'Connor, and V. Radeka. 2007. Preliminary studies of a simultaneous PET/MRI scanner based on the RatCAP small animal tomography. *Nucl. Instr. Meth. A* 571:102–105.

Yamamoto, K., K. Yamamura, K. Sato, S. Kamakura, T. Ota, H. Suzuki, and S. Ohsuka. 2007. Newly developed semiconductor detectors by Hamamatsu. Paper presented at Int. Workshop on New Photon Detectors, Kobe, Japan.

Yamamoto, S., S. Takamatsu, H. Murayama, and K. Minato. 2005. A block detector for a multislice, depth-of-interaction MR-compatible PET. *IEEE Trans. Nucl. Sci.* 52:33–37.

# 8 Current-Mode Front-End Electronics for Silicon Photomultipliers

*Francesco Corsi, Cristoforo Marzocca, and Maurizio Foresta*

## CONTENTS



## 8.1 FROM APD TO SiPM

When using solid-state photodetectors in low-light measurements, the lowest detection limit is usually imposed by the noise associated with the front-end circuit. For example, typical p–i–n (PIN) silicon photodiodes exhibit very low noise levels related essentially to the fluctuations in the leakage current of the reverse-biased p–n junction and generally referred to as parallel noise. In this case, the front-end noise, which contains essentially thermal and flicker contributions, is predominant, especially for large values of the detector capacitance. A great deal of care must be devoted to the design of the charge-sensitive preamplifier (CSA) and the shaping filter to achieve the desired performance in terms of energy resolution of the whole detection system [1, 2].

For this reason, in most low-light or even single-photon detection applications, devices with a degree of internal gain capability are usually preferred, since they allow for the release of front-end noise constraints.

Nowadays, vacuum photomultiplier tubes (PMT) are used in several applications in the fields of high-energy physics, medical imaging, and astronomy, where the incident flux of light is very low. PMTs are characterized by large gain and good timing accuracy, but they are bulky and fragile devices and require very high supply

voltages in the range of kilovolts (kV). Moreover, they are very sensitive to magnetic fields and their cost is high, especially when good performance is required. Much effort has been devoted in the past to find effective solid-state alternatives to PMTs, and essentially up to now the avalanche phenomenon has been exploited as the multiplication mechanism used to increase the intrinsic gain of semiconductor detectors.

From this perspective, avalanche photodiodes (APDs) have been successfully employed with gain factors on the order of $10^2$, thus allowing for good energy resolution at low radiation energy. The structure and the operation principle of an APD are illustrated in Figure 8.1. In the proportional mode of operation, at a given reverse-bias voltage below the breakdown threshold, each incident photon absorbed in the depletion region creates free carriers that drift toward a region where the electric field is very high, thanks to a suitable doping profile. Here, the carriers (generally electrons, since they have a greater ionization coefficient than holes) acquire sufficient kinetic energy to produce more electron-hole pairs as a result of impact ionization; thus, the carrier multiplication factor $M$ and an intrinsic gain are obtained.

An optimal value for $M$ can be found by considering that the rms value of the shot noise associated with the detector leakage is proportional to the multiplication factor $M$, whereas the thermal noise of the front-end electronics is independent of $M$. As a consequence, for low values of factor $M$, the thermal noise is dominant, as occurs in PIN detectors, whereas for high values of the same factor, the shot noise prevails and the signal-to-noise ratio degrades. A maximum signal-to-noise ratio is

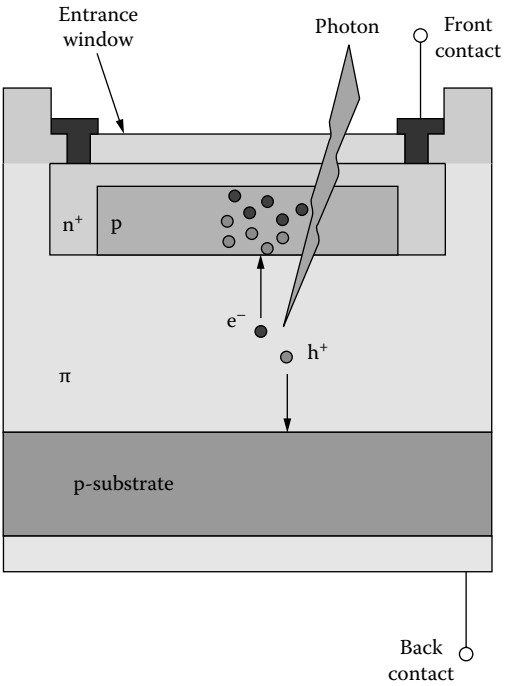



**FIGURE 8.1** Principal structures of avalanche photodiode.

thus found where the contributions from the shot noise and the thermal noise are equal, corresponding to the value of $M_{opt}$ [3]. Typically, $M_{opt}$ is found in the range between 10 and 100.

The device bias must be applied well below the breakdown voltage to avoid an increase in the ionization rate for holes, which would lead to a positive feedback effect and sustained avalanche, i.e., breakdown [4]. Since the ionization coefficient is strongly dependent on the electric field in the avalanche region, the gain is extremely sensitive to variations of the applied bias. Furthermore, the avalanche process is affected by the temperature, and therefore the intrinsic gain of the detector exhibits remarkable variations if the temperature is not accurately controlled. Finally, excess noise is introduced by the statistical nature of the avalanche process, also associated with the previously mentioned positive feedback due to holes [5, 6]. As a consequence, APDs operated in proportional mode are not suitable for single-photon detection and need accurate control of both temperature and bias voltage.

Single-photon detection can be achieved by operating an avalanche photodiode in Geiger mode. Let us consider that the APD is biased at $V_{bias}$, above the breakdown voltage $V_{br}$, for instance by means of a voltage source and a series resistor $R_q$. In this case, a large number of carriers is produced by impact ionization, and if the generation rate is higher than the extraction rate of the carriers, the electric field inside the depletion region is influenced by the growth of the population of electrons and holes and starts to decrease. The generation rate also decreases until a balanced situation is reached and a steady-state current flows in the APD, corresponding to a certain voltage drop on the series resistor. In this case, the voltage upon the APD is very close to the breakdown, and the detector operates in a similar way to a zener diode, usually employed for voltage regulation purposes.

If the series resistor is increased, the steady-state current decreases and can become lower than the minimum value needed for a self-sustained avalanche, corresponding to several tens of microamperes (μA) [7, 8]. In this case, the avalanche is extinguished, or "quenched," and the current that flows in the series resistor is $(V_{bias} - V_{br})/R_q$, since, at the moment of quenching, the voltage across the photodiode is very close to the breakdown voltage. This current flows in the photodiode capacitance $C_d$ and restores the initial charge and field distributions in the photodiode. In this "recovery" phase, characterized by the time constant $\tau_q = R_q C_d$, the voltage across the photodiode increases from the breakdown $V_{br}$ to the supply voltage $V_{bias}$. Thus, the total charge extracted from the detector corresponding to an avalanche process is

$$Q = C_d(V_{bias} - V_{br}) \tag{8.1}$$

When a photoelectron generated by an external photon absorbed in the low-field depletion region reaches the high-field region, an avalanche is triggered. This is almost immediately self-quenched and followed by the recovery phase, according to the mechanism described previously. The resulting collected charge, $Q$, is independent of the number and the energy of the incident photons and linearly related to the so-called overvoltage, $\Delta V = V_{bias} - V_{br}$. Typical gains are in the range of $10^5$–$10^6$, and therefore the thermal noise of the front-end electronics does not play a relevant role. Geiger-

mode-operated APDs are thus sensitive to single photons and are often referred to as SPADs (single-photon avalanche detectors). Unlike PMTs, they are insensitive to magnetic fields, have low bias voltage, and are compact and inexpensive. Moreover, Geiger-mode operation ensures greater stability against variations of bias voltage and temperature when compared to the proportional mode. As a consequence of the very short time constants associated with the avalanche process, excellent timing accuracy is obtained with SPAD detectors; single photoelectron timing resolutions in the range of 50 ps FWHM (full width at half maximum) can easily be achieved [9].

In the Geiger-mode APD described, the quenching of the avalanche process and the subsequent recovery are caused by the presence of the series resistor $R_q$. This simple approach is referred to as passive quenching, as opposed to the so-called active quenching, where the avalanche is turned off very quickly and the photodiode is brought back to initial conditions by means of an active circuit that senses the rising edge of the current pulse produced by the detector. The reduction of the recovery phase allows for a complete exploitation of the timing characteristics of the Geiger-mode APD [7].

The main source of noise in this kind of detector is due to Geiger discharges caused by carriers thermally generated in sensitive volume. The resulting "dark rate" is typically in the range of megahertz per square millimeter (MHz/mm$^2$) and, of course, decreases with the temperature. Another relevant source of error is due to carriers captured by deep levels in the depletion region during an avalanche and then released during the recovery phase. These carriers can retrigger an avalanche before the completion of the recovery phase, thus causing a further Geiger discharge with a total delivered charge lower than $Q$, since the discharge starts from an initial voltage lower than $V_{bias}$. This phenomenon, known as *after-pulsing*, produces current pulses of smaller amplitude following the main Geiger discharge and represents a source of error in the total amount of charge associated with a single photon detection.

Since the charge associated with the Geiger discharge is independent of the number of incident photons, a single Geiger-mode APD cannot provide information about this number. If several APDs operated in Geiger mode—realized on the same substrate and individually quenched by a series resistor—are connected in parallel, the total charge provided by the resulting device (called a *silicon photomultiplier*, SiPM, or a *multipixel photon counter*, MPPC) is proportional to the number of incident photons, provided that the probability that more than one photon hits each single photodiode is negligible [10–12]. Figure 8.2 shows the structure of a SiPM, which benefits from all the advantages of the SPAD in terms of gain, speed, compactness, and robustness, but also shows the same drawbacks, such as after-pulsing and dark rate.

The probability that a photon incident upon the surface of the detector will produce a detection event is a very important performance parameter of the SiPM, known as *photon detection efficiency* (PDE). The total PDE of the SiPM is the product of three contributions. The first is the area fill factor, i.e., the ratio between the sensitive area and the total area of the array: The main causes that limit the fill-area factor are the quenching resistor associated with each photodiode in the elementary microcell, the metal grid used to distribute the bias voltage all around the surface of the detector, and the dead space between the microcells. The second factor is the quantum efficiency, defined as the probability that an incident photon is absorbed in the depletion

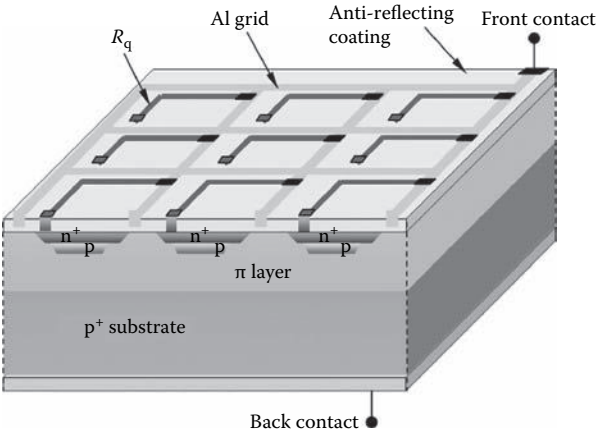



**FIGURE 8.2** Schematic structure of a silicon photomultiplier.

region of the microcell. Finally, there is the probability that a photoelectron (or hole) may start an avalanche process that is not prematurely extinguished. Measured values of the PDE can reach more than 30% for green light [12] and are typically less than 20% for blue light [11].

Another important effect that limits the accuracy and the resolution of an SiPM is optical cross talk: an avalanche process taking place in a microcell generates photons with sufficient energy to create electron-hole pairs in the depletion region of another microcell. Thus, the Geiger discharge of a microcell can trigger an avalanche in another microcell. This phenomenon can be effectively limited by adopting suitable technological solutions, such as optical isolation structures around the single microcell. Unfortunately, these structures occupy part of the total area of the detector and, as a consequence, have a negative influence on the total PDE of the detector.

Due to their excellent characteristics in terms of quantum efficiency, high gain, low bias voltage, insensitivity to magnetic fields, and timing performance, potential applications of SiPM detectors span from high-energy physics (for instance, electromagnetic calorimeters) to medical imaging (positron-emission tomography [PET], time-of-flight PET, PET/MRI [magnetic resonance imaging] systems) as well as biological applications, such as cell imaging and biodiagnostic equipment, and in astronomy (single photon counting for telescopes).

## 8.2 THE MODEL

In order to fully exploit the excellent gain and timing characteristics of the SiPM, specific front-end electronic support is needed, featuring both large bandwidth and wide dynamic range, whereas in most applications noise does not represent a major concern, due to the large gain of the detector.

Different architectures have been adopted to read out SiPM detectors in single-channel applications, ranging from voltage amplifiers (VAs) to transimpedance amplifiers (TIAs) to charge-sensitive amplifiers (CSAs), very often exploiting circuit

solutions used for PMTs. In many recently developed applications that employ large arrays of SiPM channels, compactness and reliability needs require integrated solutions for the front-end electronics.

Once the most suitable front-end (FE) architecture has been chosen for the specific application, design optimization is carried out by means of circuit simulations. For this purpose, in-depth knowledge of the behavior of the detector as a signal source and the availability of an accurate electrical model of the SiPM are of particular importance. Reliable circuit-level simulations of the detector coupled to the front-end electronics can be performed so that the main characteristics of the waveform of the achieved signals can be conveniently related to both SiPM parameters and front-end characteristics. The accuracy of the simulations clearly depends on the choice of the model parameters, which must be selected following an effective characterization procedure based on measurements from the real device.

A circuit model of the device that appears to be appropriate for most applications is shown in Figure 8.3, with $N$ representing the total number of microcells [13, 14].

This model highlights the main circuit parameters of the SiPM and, together with the corresponding parameters of the front-end (FE) electronics, allows for the description of the dynamic behavior of the SiPM+FE system. It is derived from the model of the SPAD but also includes the effects of the nonfired microcells and other parasitic effects. The model of the single microcell, apart from the diode capacitance $C_d$ and the quenching resistor $R_q$, also contains $C_q$, a small parasitic capacitor parallel to $R_q$, which works as a fast path for the charge delivered during the avalanche [7]. As a result, Equation (8.1) should be slightly modified, to take into account the contribution of $C_q$

$$Q = (C_d + C_q)\Delta V \tag{8.2}$$

Because a large metal routing, which spans the entire detector surface, is used to connect the microcells in parallel, a further parasitic capacitance, $C_g$, has been introduced in the model between the terminals of the device. For instance, if an SiPM of 1-mm$^2$ area is considered, and assuming that the metal routing grid covers 35% of

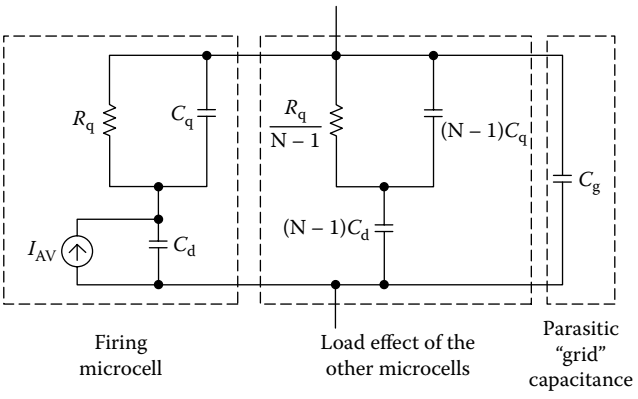



**FIGURE 8.3** Electrical model of a silicon photomultiplier.

the total surface, a value of about 11 pF can be estimated for $C_g$, considering only the contribution due to a typical metal-to-substrate capacitance per unit area of 0.03 fF/μm$^2$ while neglecting additional contributions such as those of the fringe capacitance of the metal lines, the bonding pad, etc.

Finally, the source $I_{AV}$ models the current pulse associated with the avalanche breakdown. A precise evaluation of the detection system timing performance would require an exact knowledge of the avalanche buildup to model the rise time of current pulse $I_{AV}$ accurately. On the other hand, the superposition of the elementary current pulses produced by each individually fired microcell allows us to evaluate the peak amplitude of the resulting waveform and hence of the dynamic range required by the FE. For instance, when the SiPM is coupled to a scintillation crystal, the output waveform becomes a result of the convolution between the elementary current pulses produced by the microcells and the scintillator decay law. However, in the vast majority of applications, the time constants introduced by the external circuit are much slower than those associated with the avalanche phenomenon, and so it is sufficient to model the elementary current pulse $I_{AV}$ as a simple Dirac pulse.

As far as the parameter extraction procedure is concerned, the first step is to measure the quenching resistor $R_q$. This can be carried out easily by biasing the SiPM so that the diodes composing the detector operate in the forward region. In this way, all the variations of the bias voltage on the detector are absorbed by the quenching resistors, since the voltage drop on the forward-biased diodes exhibits only negligible variations. As a consequence, the $1/R_{qtot}$ slope of the forward I-V static characteristic of the device directly provides the value of $R_q = NR_{qtot}$. Figure 8.4 shows such a characteristic for an SiPM manufactured by FBK-Irst. The linear fit of the curve is also shown, indicating that the nonlinear contribution of the diodes to the voltage drop is definitely negligible. From this curve, a slope of about 1.6 mS has been extracted, which results in a value of about 393 kΩ for $R_q$, where $N = 625$.

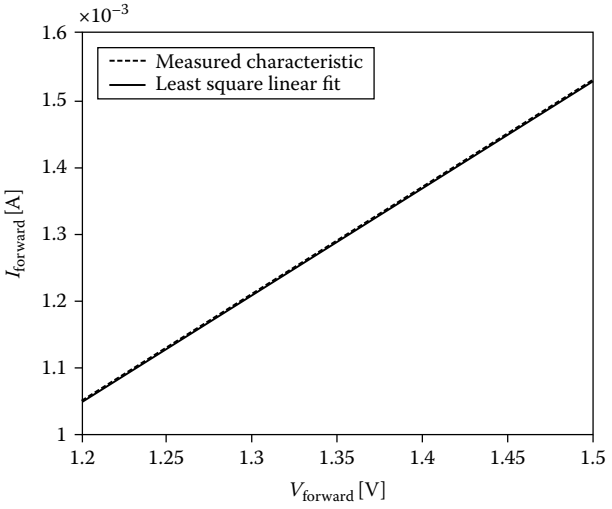



**FIGURE 8.4** Typical forward I-V characteristic of a silicon photomultiplier.

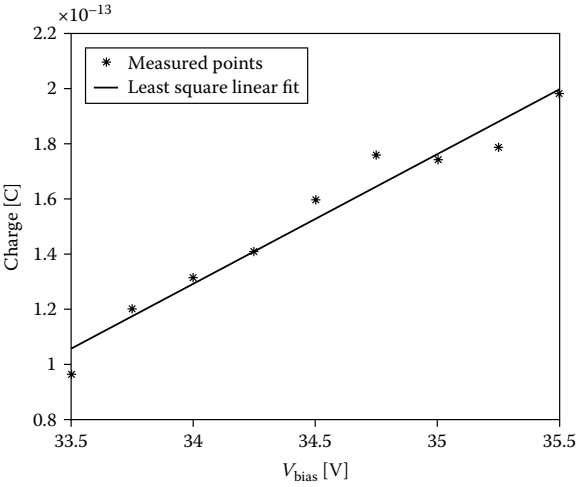



**FIGURE 8.5** Charge associated with a single dark-current pulse as a function of the bias voltage.

Equation (8.2) can be used to extract the value of both the sum $C_d + C_q$ and the breakdown voltage $V_{br}$. The charge delivered by a microcell after a Geiger discharge can be measured easily by considering single dark-count pulses read out by means of a simple front-end channel with known gain. For instance, the current delivered by the SiPM can be converted into a voltage by means of a linear resistor, and this voltage is amplified by a voltage amplifier, assembled with discrete components. Notice that the bandwidth of the amplifier does not affect the integral of the output voltage pulse; therefore, an accurate charge measurement can be carried out. This charge measurement is performed at different bias voltages $V_{bias}$ applied to the SiPM. According to Equation (8.2), the slope of the curve obtained relating $Q$ to $V_{bias}$ provides the total microcell capacitance $C_d + C_q$, whereas the $x$-axis intercept gives the breakdown voltage $V_{br}$. In Figure 8.5, such a curve is shown for the FBK-Irst device.

Finally, a capacitance-voltage (CV) plotter is used to measure the total conductance $Y_m$ and capacitance $C_m$ at the SiPM terminals, when the detector is biased just below the breakdown voltage. The instrument considers the device as a black box consisting of the parallel connection of a linear resistor and a capacitor. As a result, the conductance and the capacitance actually measured must be interpreted in the light of the SiPM model in Figure 8.3. This leads to the following expressions of $Y_m$ and $C_m$ as functions of the SiPM parameters, as reported in Equations (8.3) and (8.4),

$$Y_m = \frac{\omega^2 R_{qtot} C_{dtot}^2}{1 + \omega^2 R_{qtot}^2 C_t^2} \tag{8.3}$$

$$C_m = \frac{C_{dtot} + C_g + \omega^2 R_{qtot}^2 C_t \left( C_g C_t + C_{qtot} C_{dtot} \right)}{1 + \omega^2 R_{qtot}^2 C_t^2} \tag{8.4}$$

where $C_{dtot} = NC_d$, $C_{qtot} = NC_q$, $C_t = N(C_d + C_q)$, and $\omega$ is the frequency of the signal used by the CV plotter to perform the measurements. Using these equations, it is possible to extract the values of $C_d$, $C_q$, and $C_g$. Table 8.1 summarizes the results of the extraction procedure applied to two SiPM detectors produced by different manufacturers.

The SiPM model and the extraction procedure have been validated by coupling the detectors considered in Table 8.1 to amplifiers with known characteristics and comparing the results of SPICE (simulation program with integrated circuit emphasis) simulations of single dark pulses carried out using the extracted parameters to the measured waveforms. Very good fittings between simulated and measured data have been achieved. An example is given in Figure 8.6, which shows the measured and simulated responses to a single dark pulse of the SiPM from FBK-Irst, described

**TABLE 8.1**
**Results of the Parameter-Extraction Procedure Applied to Two SiPMs from Different Manufacturers**

| Model Parameter | SiPM FBK-Irst $N = 625$ | SiPM Photonique $N = 516$ |
|---|---|---|
| $R_q$ | 393 kΩ | 774 kΩ |
| $V_{br}$ | 31.2 V | 61 V |
| $Q$ | 175.5 fC | 127.1 fC |
| $C_d$ | 34.6 fF | 40.8 fF |
| $C_q$ | 12.2 fF | 21.2 fF |
| $C_g$ | 27.8 pF | 18.1 pF |

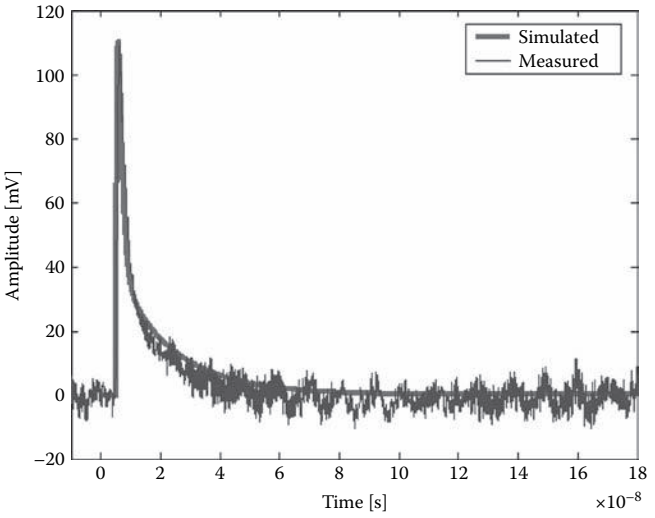



**FIGURE 8.6** Measured and simulated dark-pulse responses for the FBK-Irst SiPM coupled to a voltage amplifier.

in Table 8.1, coupled to a fast voltage amplifier with an input resistance of $R_S \cong 50\ \Omega$, bandwidth of 380 MHz, and gain equal to 140.

## 8.3 CURRENT-MODE FRONT-END APPROACH

Among the different front-end architectures that can be considered to read out an SiPM, the most straightforward solution is represented by the charge-sensitive amplifier (CSA), in which the charge delivered by an event is collected on a feedback capacitor, as shown in Figure 8.7.

This configuration is able to guarantee the best noise performance and represents the standard solution for radiation detector readout, but in the case of SiPM, it poses a number of issues. Because the charge delivered by the detector in response to an event is very large, a large integration capacitance $C_f$ is needed to keep the output voltage within the allowed dynamic range imposed by the power supply. This limitation becomes quite severe when deep-submicron technology, characterized by very low power supply voltage, is used. For instance, considering a typical gain value of $10^6$, if 300 microcells of the SiPM are hit, a total charge of about 48 pC must be integrated on the feedback capacitance of the CSA stage. If a standard CMOS (complementary metal-oxide semiconductor) 0.35-μm technology, featuring a typical power supply voltage of 3.3 V, is used to implement the front end, the maximum allowed voltage swing at the CSA output does not exceed 3 V. Thus a feedback capacitance of about 16 pF is needed, which is impractical in a front-end chip containing several tens of channels, because of the excessive silicon area occupancy. The situation worsens in cases where a 0.25-μm (or even less) technology is used and/or the total number of microcells contained in the SiPM is increased, as is often required by several applications.

Moreover, the preamplifier stage included in the feedback loop of the CSA should drive a heavy capacitive load given by the series connection of the feedback and the detector capacitances. To guarantee good speed performance, a suitable output stage, able to drive this capacitive load, must be arranged. This would entail a significant increase in the power consumption for the front-end channel.

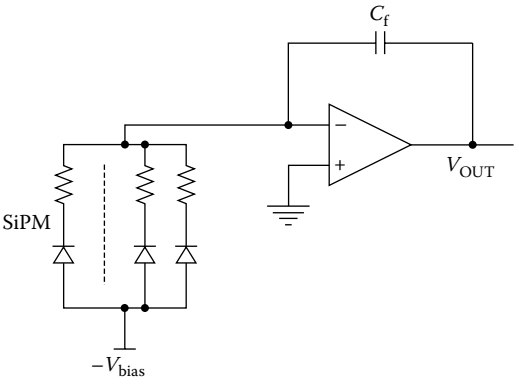



**FIGURE 8.7** SiPM readout by a charge-sensitive preamplifier.

In previous realizations based on the CSA approach, the detector was AC-coupled to the front end by means of an integrated capacitance, which forms a capacitive charge divider with the detector capacitance (several tens of picofarads), thus reducing the total charge that must be integrated onto the feedback capacitance [15]. On the one hand the detector is not connected to a virtual ground, thus it experiences variations of the bias voltage, which depend on the signal amplitude, while on the other hand the high gain of the detector is not fully exploited for the extraction of the timing information.

Very often, the readout approach consists of a current-to-voltage conversion of the signal provided by the detector through a linear resistor $R_S$, followed by a fast voltage amplifier to achieve the desired signal level. This arrangement has been used frequently for the characterization of SiPMs. First, the resistor value should be small enough to avoid fluctuations of the detector bias voltage under signal and the introduction of low-frequency poles at the amplifier input. Timing can be made faster than in the CSA case, but the integration of the output signal, needed to extract the value of the charge (and thus the energy associated with an event), requires a further voltage-to-current reconversion, with a suitable scaling-down factor with respect to the detector current in order to avoid the dynamic-range problems mentioned previously. Note that the overall operation of such a system would be equivalent to a current amplifier.

A transimpedance amplifier would suffer more or less from the same problems, since the output signal is a voltage, as in the case of the voltage amplifier. A further issue to be dealt with is the stability of the feedback loop.

On the basis of the previous considerations, an interesting front-end architecture can be adopted based on the use of a current buffer characterized by a very low input impedance, as depicted in Figure 8.8.

The output signal of this preamplifier is a replica of the detector current at very high output impedance. This approach is very flexible in a CMOS implementation, since the output current can be replicated easily with different scaling factors by means of multibranch current mirrors. The circuit is inherently fast, and the current-mode operation enhances the dynamic range, since it does not suffer from possible voltage limitations due to deep-submicron implementation. A possible drawback of the current-mode approach may be an increase in the electronic noise level but, as mentioned previously, this does not represent a major concern due to the large gain of the detector.

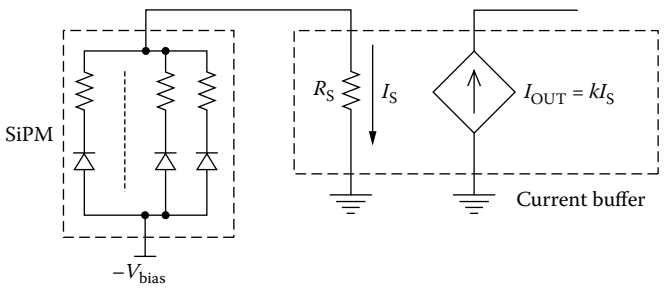



**FIGURE 8.8** SiPM readout by a current buffer.

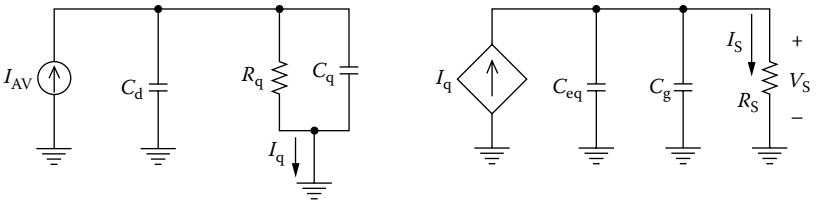



**FIGURE 8.9** SiPM coupled to the input impedance of the front-end electronics $R_s$: simplified circuit.

It is interesting to exploit the previously defined model to understand how the parameters of the SiPM and the input resistance of the electronics $R_S$ influence the resulting waveform of the electrical variables at the input of the preamplifier. If the SiPM model shown in Figure 8.3 is loaded with the resistor $R_S$, a qualitative study of the resulting circuit can be carried out with reference to the simplified diagram shown in Figure 8.9, which is a good approximation of the complete circuit, provided that $R_S$ is much lower than $R_{qtot}$, as usually happens. This has been demonstrated by means of SPICE simulations, which show that the responses of both the complete and simplified circuits to short current pulses are almost identical in cases of $R_S << R_{qtot}$.

In the simplified circuit, $C_{eq}$ represents the series connection between $(N-1)C_d$ and $(N-1)C_q$ of the model in Figure 8.3. Simple hand calculations provide the following expression for the input voltage $V_{IN}(t)$ in response to an input current pulse of area $Q$:

$$V_{IN}(t) \cong \frac{QR_S}{\tau_r - \tau_{IN}}\left(\frac{\tau_q - \tau_{IN}}{\tau_{IN}}\exp(-\frac{t}{\tau_{IN}}) + \frac{\tau_r - \tau_q}{\tau_r}\exp(-\frac{t}{\tau_r})\right) \tag{8.5}$$

where $\tau_{IN} = R_S(C_g + C_{eq})$, $\tau_r = R_q(C_d + C_q)$, and $\tau_q = R_qC_q$. Equation (8.5) accounts for the presence of two time constants in $V_{IN}(t)$. The first is $\tau_{IN}$, associated with the input impedance of the front end and to the total capacitance at the terminal of the detector, whereas the second, $\tau_r$, is the recovery time of the microcells. Notice that the peak value of $V_{IN}$ is independent of $R_S$, since the charge $Q$, released very quickly by the avalanche phenomenon, is shared between the capacitor $C_d$ and the series connection of $C_q$ and $C_{tot} = C_{eq} + C_g$, almost equal to $C_q$ itself. The resulting initial value $V_{IN}(0) = V_{INMAX}$ is thus equal to the fraction of the charge collected by $C_q$, called $Q_{IN}$, divided by $C_{tot}$

$$Q_{IN} \cong Q\frac{C_q}{C_q + C_d} \quad V_{IN}(0) \cong \frac{Q_{IN}}{C_{tot}} \tag{8.6}$$

This value is the same as that provided by Equation (8.5). As a consequence, the peak value of the current delivered by the SiPM coupled to $R_S$ is $V_{INMAX}/R_S$, which increases with decreasing values of $R_S$. The rise time of $V_{IN}(t)$ depends only on the time constants associated with the avalanche process, which are assumed to be much

faster than $\tau_{IN}$ and $\tau_r$. On the other hand, the rise time and the peak value of $I_{OUT}$ in Figure 8.8 are strongly affected by the bandwidth of the current buffer used, especially for low values of $R_S$. In the latter case, only a wideband amplifier will be able to reproduce the fast time constant $\tau_{IN}$ of $V_{IN}(t)$ at its output. Conversely, the response of an amplifier that is slow compared to $\tau_{IN}$ will only exhibit the amplifier rise time and the recovery time constant $\tau_r$.

Current feedback can be conveniently exploited to improve the performance of an input current buffer designed using standard CMOS technology. In Figure 8.10, the structure of a possible implementation of the circuit is shown.

The stage is essentially a current follower. The common gate transistor $M_1$ is closed in a feedback loop, built around the common source $M_2$ and the diode-connected MOSFET (metal-oxide semiconductor field-effect transistor) $M_4$. The negative feedback decreases the input impedance of the buffer and enhances its frequency response by a factor that depends on the loop gain $T$, equal to $g_{m2}/g_{m4}$. The locations of the open loop poles are the following:

$$\omega_1 = g_{m1}/(C_g + C_{eq}); \quad \omega_2 = g_{m4}/C_1 \tag{8.7}$$

where $C_1$ is the total capacitance, which loads the drain of $M_1$. The value of the open-loop gain has to be chosen so as to guarantee the stability of the circuit with the given constraints in terms of maximum current consumption. Generally, a low value of $T$ arises (around 10), due to the fact that the dominant open-loop pole $\omega_1$ and the second pole $\omega_2$ are not too far apart: If $g_{m4}$ is decreased to increase $T$, a more unsatisfactory situation is achieved for stability, since $\omega_2$ slows down. On the other hand, if $g_{m2}$ is increased by increasing the size of $M_2$, $C_1$ also increases, and once again a slower $\omega_2$ is obtained. The desired ratio between $g_{m2}$ and $g_{m4}$ was achieved by adjusting the bias currents of $M_2$ and $M_4$ by means of the current source $M_3$.

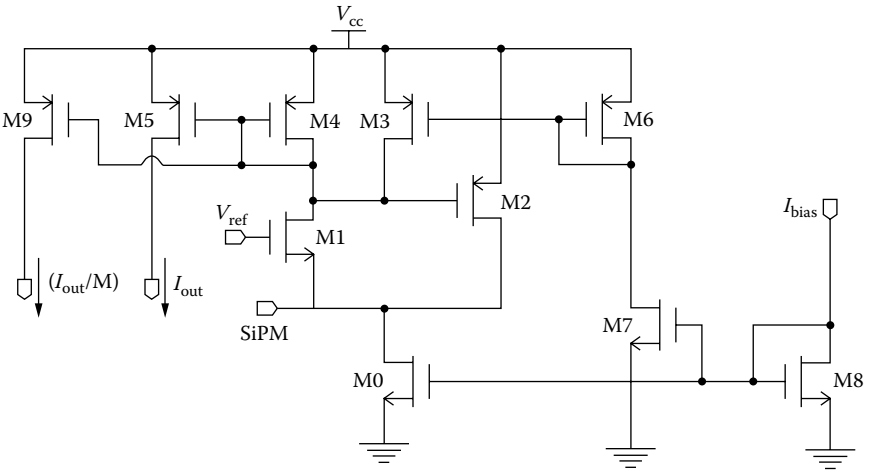



**FIGURE 8.10** CMOS implementation of the input current buffer.

**TABLE 8.2**
**Main Performance Parameters of the Designed Current Buffer**

| Performance Parameter | Value |
|---|---|
| Small signal bandwidth | 250 MHz |
| Total current consumption | 800 μA |
| Rise time of the output waveform (single microcell) | 400 ps |
| Input impedance | 17 Ω |
| Linearity dynamic range | 50 pC |
| Power supply | 3.3 V |
| Range allowed for $V_{ref}$ | 1–2 V |

Because the SiPM is DC-coupled to the current buffer, an interesting feature of this circuit is the ability of a continuous variation of the SiPM bias voltage, which can be accomplished by varying the value of $V_{ref}$ [16]. This can be useful in many applications to fine-tune the gain of the detector, which varies by about 3% for every 100 mV of bias variation [17]. Table 8.2 summarizes the main features of a current buffer designed with standard 0.35-μm CMOS technology.

Note that if a more advanced CMOS technology is used for the implementation of the circuit, the open-loop poles will be placed at higher frequencies, since higher transconductance values can be achieved with the same bias currents. Furthermore, the relative distance between the dominant poles will increase, thanks to the smaller value of the $C_1$ capacitance. As a consequence, higher loop gain values could be chosen without stability problems, thus achieving better performance in terms of both input impedance and bandwidth.

As stated previously, another relevant feature of the current buffer is flexibility: The output current can be replicated easily with different scaling factors, as shown in Figure 8.10. Consequently, different signal paths can be implemented to extract the desired information from the current pulse generated by the detector. The complete front-end architecture is shown in Figure 8.11.

In a first "slow" path, an output of the input current buffer, suitably scaled down by factor $M$:1 to cope with the dynamic-range limitations discussed previously, is sent at the input of an integrator, which provides a voltage signal with an amplitude proportional to the total charge contained in the detector current pulse. A great deal of care must be taken in the choice of the integrator dumping time constant $\tau_f = R_f C_f$, which cannot be too large, so as to avoid pileup problems and the integration of a large number of dark pulses. On the other hand, a dumping time constant that is too short with respect to the decay time constant of the scintillator used in the detection system may cause large errors in the evaluation of the total number of photons delivered by the event. In many cases, a time constant digitally programmable with a small number of bits can be an appropriate solution. Furthermore, the gain of the integrator can be made programmable to fit different applications or detectors, for instance by changing the feedback capacitance of the integrator or the scaling factor of the output current mirror in the input buffer. The first solution is preferable in our example, since a variation of the gain of a current mirror can be carried out by

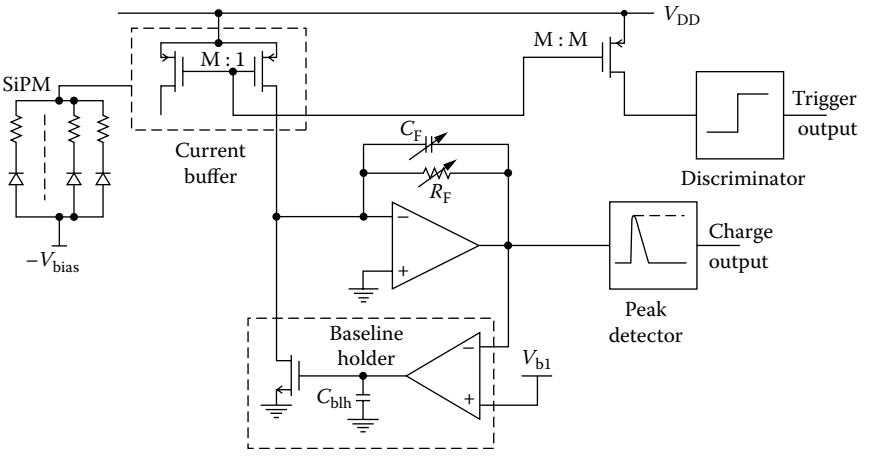



**FIGURE 8.11** Complete architecture of the front end for SiPM detectors.

increasing or decreasing the equivalent width of the output MOSFET of the mirror, and this causes variations of the capacitance $C_i$ in the current buffer. As a consequence, changes occur in the circuit bandwidth that affect the rise time and the peak of the output current, as has already been mentioned.

Another relevant issue is the control of the DC output voltage of the integrator, which must be set at a well-defined baseline level $V_{bl}$. For this purpose, the integrator is enclosed in a very slow feedback loop that compensates for the DC component of the input current but is ineffective as far as fast signals are concerned. This can be done by inserting a very-low-frequency pole into the feedback path of the loop, as graphically illustrated in the baseline holder section in Figure 8.11. When the rate of events increases, the input current pulse train starts to make a significant contribution to the DC level at the input of the feedback system, which tends to remove this additional DC component from the output pulse train. As a result, a shift of the output pulses baseline at lower levels is observed if the system is perfectly linear. Excessive variations of the baseline level as a function of the event rate can be avoided by suitably exploiting the unavoidable nonlinearity of the circuit used in the feedback path, in order to decrease the loop gain in the presence of signal pulses [18].

The information relating to the energy of incoming events is associated with the peak voltage pulse value at the output of the integrator. A peak detector circuit can be exploited to hold the peak value as long as is necessary to allow for analog-to-digital conversion. A possible implementation of this important building block is shown in principle in Figure 8.12 and is based on a feedback loop containing a pMOS current mirror $M_1 - M_2$, which works as a rectifying element [19]. In fact, as long as the input voltage increases, the mirror is able to charge the "hold" capacitor $C$, so that the output voltage follows the input. When the input pulse reaches its peak value, the current in the pMOS mirror reverts to zero, the charge stored on the capacitor is frozen, and the feedback loop is broken. The operational transconductance amplifier (OTA), used as a gain stage in the feedback loop, goes into saturation and switches

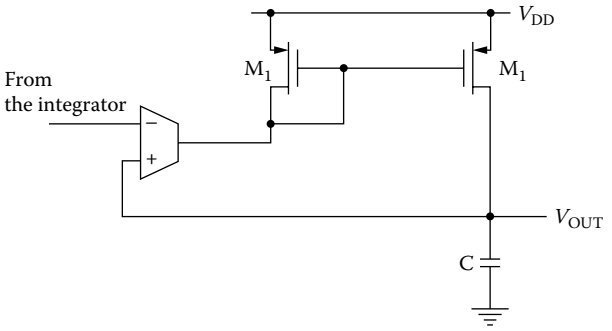



**FIGURE 8.12** A CMOS peak detector.

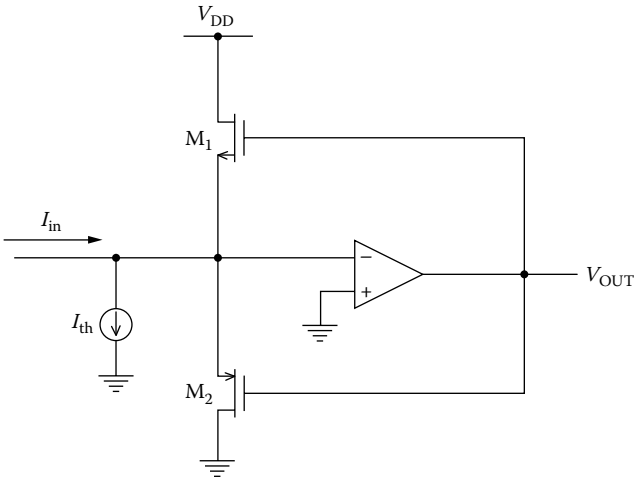



**FIGURE 8.13** The current discriminator principle in diagram form.

off the transistors of the mirror, so that the output voltage does not undergo further change and reproduces the peak of the input pulse.

A second "fast" path for the signal is needed to extract the trigger signal, which marks the arrival time of an event with its fast transition. A further output branch of the input current buffer is exploited for this purpose, without any need to scale down the current. The buffered replica of the detector current is applied to the input of a current discriminator, which produces the trigger signal.

A possible implementation of this circuit, already used in several applications [20], is shown in Figure 8.13. When the input current $I_{in}$ is smaller than the threshold $I_{th}$, the nMOS $M_1$ is on and carries the current $I_{th} - I_{in}$, thus the output voltage of the inverting amplifier is sufficiently high to keep the $V_{GS}$ of $M_1$ beyond the threshold voltage. As soon as the input current rises above $I_{th}$, the difference $I_{th} - I_{in}$ becomes negative, $M_1$ switches off, and the pMOS $M_2$ turns on; thus the output voltage of the amplifier quickly decreases. Suitable level-shifting circuitry, inserted between the gates of $M_1$ and $M_2$ but not shown in Figure 8.13, makes the discriminator faster. The output of the

inverting amplifier is buffered by means of cascaded CMOS inverters used as open-loop amplifiers, which provide a clean output pulse at full voltage swing. The discriminator threshold is easily adjusted by varying the current $I_{th}$, which may be provided by a digital-to-analog converter (DAC) driven by the desired configuration. Timing accuracies in the range of several tens of picoseconds have been achieved with this circuit at low thresholds and full-range input signals. The spread of the delay in the discriminator response to an input current pulse is in the range of several hundred picoseconds, obtained by varying both the threshold and the amplitude of the input current pulse.

## 8.4 EXAMPLE OF MULTICHANNEL ARCHITECTURE

The organization of a multichannel front-end architecture is strictly dependent on the readout strategies that must be implemented (serial, sparse, etc.), on the resources that can be accommodated on a chip, and on the amount of data that must be exchanged with the external readout system. These features are, in turn, strictly related to the requirements of the specific target application. As an example, a description follows of the simple architecture of an eight-channel ASIC developed for the example of a PET scanner prototype.

Figure 8.14 shows a block schematic diagram of the chip. The eight analog channels have been designed according to the structure shown in Figure 8.11: As previously discussed, each channel provides an analog output signal proportional to the number of SiPM microcells stimulated by the event as well as a fast trigger signal that is useful in obtaining an accurate time stamp.

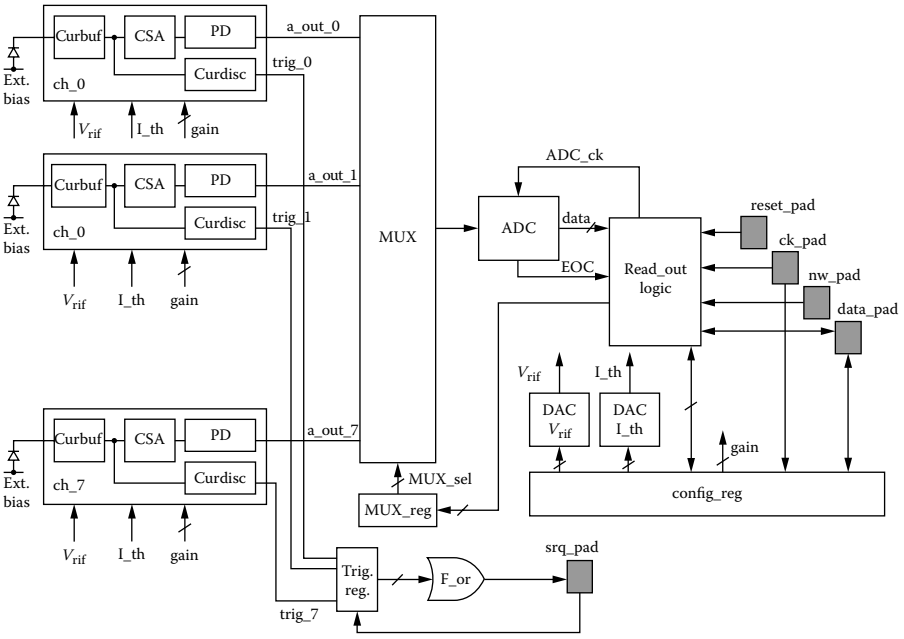



**FIGURE 8.14** Example of architecture of an eight-channel ASIC for SiPM detectors.

To cope with the high rate of valid coincidence events, the chip includes an embedded SAR ADC that converts the output of the 8-in 1-out multiplexer into an 8-bit digital word.

A configuration register contains the information for the two DACs used to adjust the bias voltage (8-bit) applied to the SiPM and set the threshold of the current discriminator (4-bit). The readout logic controls the write and read operations in the configuration register as well as the tracking and hold modes of the peak detectors.

Two data-acquisition modes are implemented: a serial readout, in which all the channel outputs are sequentially converted into digital and sent to the external bus, and a sparse one, in which only those channels interested by a hit and found over threshold are converted and sent to the output.

The trigger signals produced by either channel of the chip are stored in a trigger register to mark the channels over threshold and are sent to a fast OR gate. The resulting signal is used as a trigger for the external readout electronics, implemented for instance by means of a field-programmable gate array (FPGA). When this signal is activated, it works as a service request toward the FPGA, which puts the time stamp to the event and starts the readout procedure and data acquisition, according to the selected readout mode.

Extending this architecture to a greater number of channels is fairly straightforward. The length of the metal interconnections and the associated capacitive loads on the outputs of the individual channels call for the insertion of suitable buffers to preserve the speed of operation. This, in turn, implies an increased power absorption per channel that must fulfill the requirements of the specific application.

The decision to include an onboard ADC allows us to simplify the structure of the whole detection system as long as its conversion speed is sufficient to cope with the expected rate of events. This is in contrast to the solution of adopting an external ADC, which would imply handling an analog signal at the output of the front-end chip, with a consequent need for further buffer stages and additional power requirements. The resolution of the ADC must be chosen on the basis of the number of microcells of the SiPM. For example, in our PET application, an 8-bit resolution proved to be sufficient, as we used SiPMs with 625 microcells each, also taking into account the effects of dark count and after-pulse signals. However, the use of deep-submicron CMOS technologies would allow us to attain better speed performances, thus making the choice of an embedded ADC feasible even in the case of high event rates.

Moreover, a new technique has recently been proposed [21] in cases where there are a large number of channels on each chip. This is based on a "de-randomization" of events that are temporarily stored in analog memory before being sent to the ADC. In this way, it is possible to fully exploit the speed performance of the ADC, which is only required to match the average rate of events.

Further readout-acquisition modes such as "winner takes all," "first takes all," "sparse neighborhood," etc., may be required, depending on the specific application. In addition, in some applications, summing the electrical charges of a set of adjacent channels allows for an estimate of the energy of the incident photon.

The architecture described previously is self-triggered, since it extracts the trigger signal from the fast path of the front end. However, in some cases, it may be convenient to control the acquisition by means of an external trigger. In other cases, such

as, for example, PET systems, the acquisition needs to start only when a coincidence signal, generated within a very short time window [22], acknowledges the validity of the trigger signal.

## 8.5 SUMMARY

This chapter has presented a novel front-end architecture suitable for detectors characterized by high speed, a relatively large internal capacity value, and high gain. The circuit relies on the use of a simple current-follower stage whose performance is enhanced through the use of a current-feedback network. Thanks to this circuit arrangement, the full dynamic range of the current pulse generated by the detector is exploited to extract the trigger signal. For most applications, the timing accuracy provided by the circuit is adequate, even with the use of 0.35-μm CMOS technology.

Another advantage of the current-mode approach is that it is much less affected by the dynamic-range limitations of a voltage-mode approach, even with the use of a more advanced deep-submicron technology that entails better time performances due to reduced internal capacitances and to greater transconductance values without increasing the bias currents.

In particular, the lower values of the internal capacitances cause the two dominant poles of the front end to remain a fair distance from each other, thus resulting in a better phase margin. Moreover, a lower input impedance is also obtained thanks to both the greater transconductance of the input transistor and to the higher loop gain that can be achieved without stability problems.

The use of BiCMOS or Si-Ge technologies would also allow us to implement a better current follower in terms of both input impedance and speed by exploiting the availability of good-quality bipolar transistors. Resorting to these more costly solutions could prove to be advantageous in applications where particularly good time performances are required.

## REFERENCES

1. Chang, Z. Y., and W. Sansen. 1991. *Low-noise wide-band amplifiers in bipolar and CMOS technologies*. Dordrecht: Kluwer Academic, chap. 5.
2. Gatti, E., and P. Manfredi. 1986. Processing the signal from solid-state detectors in elementary-particle physics. *Rivista del Nuovo Cimento* 9, no. 1.
3. Hamamatsu Photodiode. 2009. Technical information: Characteristics and use. www.hamamatsu.com.
4. Ruegg, H. W. 1967. An optimized avalanche photodiode. *IEEE Transactions on Electron Devices* 14:239–251.
5. Spieler, H. 2005. *Semiconductor detector systems*. Oxford: Oxford University Press.
6. Knoll, G. F. 2000. *Radiation detection and measurement*. New York: John Wiley & Sons.
7. Cova, S., M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. 1996. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied Optics* 35:1956–1976.
8. McIntyre, R. J. 1972. The distribution of gains in uniformly multiplying avalanche photodiodes: Theory. *IEEE Transactions on Electron Devices* 19:703–713.

9. Giudice, A., M. Ghioni, R. Biasi, F. Zappa, S. Cova, P. Maccagnani, and A. Gulinatti. 2007. High-rate photon counting and picosecond timing with silicon-SPAD based compact detector modules. *Journal of Modern Optics* 54:225–237.
10. Bondarenko, G., P. Buzhan, B. Dolgoshein, V. Golovin, E. Guschin, A. Ilyin, V. Kaplin, A. Karakash, R. Klanner, V. Pokachalov, E. Popova, and K. Smirnov. 2000. Limited Geiger-mode microcell silicon photodiode: New results. *Nuclear Instruments and Methods in Physics Research A* 442:187–192.
11. Buzhan, P., B. Dolgoshein, L. Filatov, A. Ilyin, V. Kantzerov, V. Kaplin, A. Karakash, F. Kayumov, S. Klemin, E. Popova, and S. Smirnov. 2003. Silicon photomultiplier and its possible applications. *Nuclear Instruments and Methods in Physics Research A* 504: 48–52.
12. Golovin, V., and V. Saveliev. 2004. Novel type of avalanche photodetector with Geiger mode operation. *Nuclear Instruments and Methods in Physics Research A* 518: 560–564.
13. Corsi, F., A. Dragone, C. Marzocca, A. Del Guerra, P. Delizia, N. Dinu, C. Piemonte, M. Boscardin, and G. F. Dalla Betta. 2007. Modelling a silicon photomultiplier (SiPM) as a signal source for optimum front-end design. *Nuclear Instruments and Methods in Physics Research A* 572:416–418.
14. Corsi, F., C. Marzocca, A. Perrotta, A. Dragone, M. Foresta, A. Del Guerra, S. Marcatili, G. Llosa, G. Collazzuol, G. F. Dalla Betta, N. Dinu, C. Piemonte, G. U. Pignatel, and G. Levi. 2006. Electrical characterization of silicon photo-multiplier detectors for optimal front-end design. *IEEE Nuclear Science Symposium Conference Record* 2:1276–1280.
15. Bouchel, M., F. Dulucq, J. Fleury, C. de La Taille, G. Martin-Chassard, and L. Raux. 2007. SPIROC (SiPM integrated read-out chip): Dedicated very front-end electronics for an ILC prototype hadronic calorimeter with SiPM read-out. Paper presented at IEEE Nuclear Science Symposium Conference Record, Honolulu.
16. Corsi, F., M. Foresta, C. Marzocca, G. Matarrese, and A. Del Guerra. 2007. Current-mode front-end electronics for silicon photo-multiplier detectors. Paper presented at IEEE 2nd International Workshop on Advances in Sensors and Interfaces (IWASI'07), Bari, Italy.
17. Buzhan, P., B. Dolgoshein, A. Ilyin, V. Kantserov, V. Kaplin, A. Karakash, A. Pleshko, E. Popova, S. Smirnov, Yu. Volkov, L. Filatov, S. Klemin, and F. Kayumov. 2001. An advanced study of silicon photomultipliers. *ICFA Instrumentation Bulletin*, Fall.
18. De Geronimo, G., P. O'Connor, and J. Grosholz. 2000. A CMOS baseline holder (BLH) for readout ASICs. *IEEE Transactions on Nuclear Science* 47:818–822.
19. De Geronimo, G., P. O'Connor, and A. Kandasamy. 2002. Analog CMOS peak detect and hold circuits, Part 1: Analysis of the classical configuration. *Nuclear Instruments and Methods in Physics Research A* 484:533–543.
20. Heijne, E. H. M., M. Campbell, W. Snoeys, et al. 1996. LHC1: A semiconductor pixel detector readout chip with internal, tunable delay providing a binary pattern of selected events. *Nuclear Instruments and Methods in Physics Research A* 383:55–63.
21. Dragone, A., G. De Geronimo, J. Fried, A. Kandasamy, P. O'Connor, and E. Vernon. 2005. The PDD ASIC: Highly efficient energy and timing extraction for high-rate applications. *IEEE Nuclear Science Symposium Conference Record* 2:914–918.
22. Dragone, A., F. Corsi, C. Marzocca, P. Losito, D. Pasqua, E. Nappi, R. De Leo, J. Séguinot, A. Braem, E. Chesi, C. Joram, P. Weilhammer, F. Garibaldi, and H. Zaidi. 2006. An event driven read-out system for a novel PET scanner with Compton enhanced 3-D gamma reconstruction. *IEEE Transactions on Nuclear Science* 53: 1156–1161.

# 9 Integrated Charge-Measuring Systems for Radiation Detectors in CMOS Technologies

*Angelo Rivetti*

## CONTENTS



## 9.1 INTRODUCTION

In the past 25 years, significant resources have been invested in the design and implementation of complex integrated circuits for the readout of semiconductor radiation detectors. This effort was triggered by the unique capabilities offered by silicon sensors in the detection of short-lived particles in high-energy physics experiments. The

possibility of patterning a silicon surface into many independent detecting elements with a typical pitch of 50–100 μm was the key to achieve the spatial resolution necessary, for instance, to study charmed-particle decay. It was soon recognized that such a dense detector could not be read out by conventional electronics built from discrete components, and that more-compact solutions had to be developed. The use of silicon sensors read out by an application-specific integrated circuit (ASIC) was pioneered in the silicon vertex detector of the Mark II experiment at the Stanford Linear Accelerator Center (SLAC) (Adolphsen et al. 1992) and quickly became a standard practice in all high-energy physics experiments.

The physics goals of the Large Hadron Collider (LHC), now commissioned at the European Laboratory for Particle Physics (CERN), near Geneva, pushed the requirements for integrated front-end electronics in term of speed, power consumption, data handling, and radiation tolerance to unprecedented levels, motivating further research efforts. Several technologies were considered, but in the end, commercial deep-submicron CMOS (complementary metal-oxide semiconductor) emerged as the winning option. Indeed, a dedicated R&D program at CERN demonstrated that, if some layout practices were applied, integrated circuits fabricated in quarter-micron processes could withstand radiation doses well beyond what was demanded by the LHC detectors (Snoeys et al. 2000). Primarily conceived for the implementation of low-power digital electronics, modern CMOS technologies also offer good analog performance and allow the design of sophisticated mixed-mode integrated circuits incorporating key functions like amplification, filtering, and digitization on the same chip.

Along the years, high-density silicon detectors started to be exploited outside the high-energy physics domain for which they were originally conceived. X-ray detection systems for protein crystallography, digital radiography, and other applications based on silicon microstrip or pixel detectors and their associated front-end electronics were demonstrated. In the meanwhile, the benefits of segmentation (increased spatial resolution, smaller parasitic capacitance) were also applied to other sensors. Multianode photomultiplier tubes (MAPTs) with 16–64 independent anodes and pixelated avalanche photodiodes (APDs) with a comparable number of channels are available and are increasingly employed in different applications, including positron emission tomography (PET). Multichannel integrated front-end circuits are hence necessary if large systems based on these highly segmented sensors have to be assembled. As an example, a recently built scanner for positron emission mammography (PEM) makes use of 64 integrated circuits with 192 channels per chip to read out 12,288 APD channels (Trindade et al. 2008).

CMOS technologies have become the preferred options for the implementation of high-density front-end circuits for radiation sensors. Therefore, the discussion in this chapter is centered on CMOS designs. The emphasis is put on those architectures that are not limited to the detection of a hit, but measure the amount of charge and thus the energy released in the sensor by an ionizing event. Energy information is, in, fact a fundamental tool in many areas, from spectroscopy and calorimetry in nuclear and high-energy physics to photon detection in PET systems, to name just a few. The circuits described here combine simpler building blocks, like the front-end amplifiers and the analog-to-digital converters (ADCs) discussed elsewhere is this book, to implement multichannel readout systems that process the charge released

by the radiation sensor performing amplification, filtering, data sparsification, and digitization all on the same chip.

The design of such integrated circuits is a lengthy, risky, and expensive task, which starts with a proper definition of circuit specifications and the selection of a suitable architecture. Therefore, Section 9.2 of the chapter discusses the key specifications facing the chip designer and their impact on the overall performance. Front-end chips for detectors are often designed by small groups in research labs that are chronically understaffed. Detailed analysis and clear understanding of the system requirements are therefore extremely important if one wants to invest the scarce resources in the proper direction and achieve satisfactory results. Section 9.3 provides a review of the major architectures that have emerged along the years. In the past two decades, several integrated front-end systems have been developed for different applications. The examples reported in Section 9.3 have been selected just for illustrative purposes and are only a very small subset of the many interesting designs the reader can find in the open literature. As it will be apparent from Section 9.3, modern front-end systems for radiation detectors often combine—on the same silicon substrate—sensitive analog circuits with complex digital functions. The chapter ends with Section 9.4, which introduces the major issues that have to be addressed to make this coexistence successful.

## 9.2 SPECIFYING A CHARGE-MEASURING INTEGRATED CIRCUIT

Figure 9.1 shows a straightforward implementation of a charge-measuring system. In each channel, the signal coming from the sensor is amplified and filtered by a front-

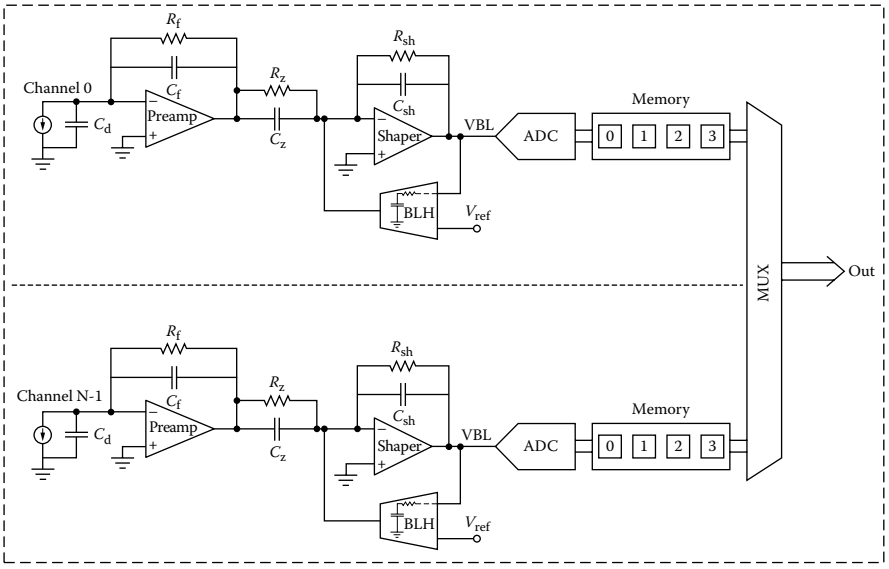



**FIGURE 9.1** Generic block diagram of a multichannel integrated circuit for charge measurement.

end amplifier-shaper and digitized by a flash ADC. The result of the conversion is stored in local memory and sent out for further processing. A multiplexer is used to reduce the number of output lines. Similar schemes are often employed to implement sensor test setup with off-the-shelf equipment. The major aspects involved in the realization of such a chain in the form of an integrated circuit are discussed in Section 9.3. Here, we will use it as an aid to develop our discussion on system specifications on a more concrete basis. Several factors enter in the design of a charge-measuring ASIC: sensor capacitance, sensor dark current, noise, dynamic range and linearity, event rate and efficiency, and power consumption.

### 9.2.1 Sensor Capacitance

As shown in Figure 9.1, the individual detecting element is often depicted as a current source in parallel to a lumped capacitor (here labeled $C_d$) representing all the parasitic capacitance seen by the preamplifier input node. Depending on the detector type, the value of $C_d$ may change by more than three orders of magnitude: from the hundreds of femtofarads typical of silicon pixel detectors to the hundreds of picofarads associated with the large-area APDs employed in calorimetry applications. It is well known that the input capacitance plays a fundamental role in setting the ultimate noise figure of the analog front end. However, it also has important system-level implications for the performance of a highly integrated front-end system. We can appreciate better this point with the help of Figure 9.2.

Here we see two complementary versions of an amplifier topology frequently adopted in the very first stage of a front-end chain: a charge integrator built by placing a capacitor in the feedback path of a single-ended cascode stage. In Figure 9.2(a), the input transistor is a pMOS, and its source is tied to the positive power supply. The transistor is biased with a high impedance, constant current source that is obtained via a cascode current mirror. Thus any modulation of the source potential will appear on the gate (note that constant bias current implies constant gate-source voltage) and will be transferred to the output approximately amplified by the factor $C_d/C_f$. The complementary configuration of Figure 9.2(b), where the input transistor is an nMOS with the source connected to ground, is better, since in this case the same modulation will manifest on both plates of $C_d$ and will appear unamplified at the output node. However, this in not in general true, since the ground to which the bottom plate of $C_d$ is connected is a "small signal ground," i.e., a constant potential needed for proper biasing of the sensor. Therefore, its DC value can be significantly different from the amplifier's ground or power supply, and the two cannot always be physically shorted. In a typical case, the bottom plate of $C_d$ is the back-plane of the detector.* This is common to all the sensing elements integrated on the same substrate and is connected to detector high-voltage bias. A link between this voltage

* In most multichannel sensors, the capacitance between one sensing element and its neighbors also gives an important contribution to the total value of $C_d$ and cannot be neglected. However, the other plate of this capacitance is connected to the input of another amplifying channel that will share the same power supply.

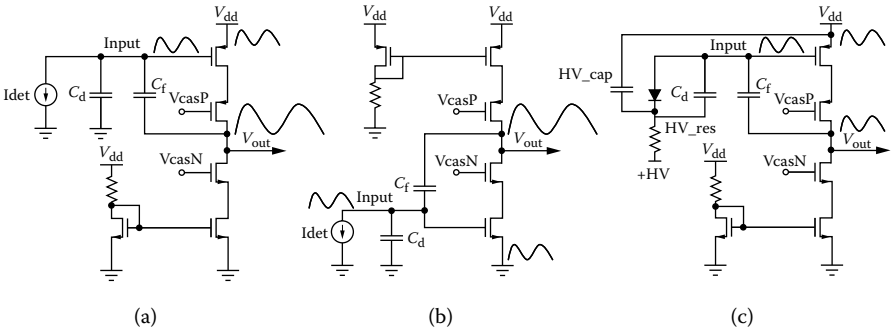



**FIGURE 9.2** Effect of the detector capacitance on the amplifier response to a disturbance appearing on the sensitive node of the input transistor, which is biased by a cascoded current source. (a) The case with pMOS input transistor; (b) the case with nMOS input; (c) the proper connection of a bypass capacitor when an amplifier with pMOS input is used to read out a sensor biased by a positive high-voltage supply.

and the amplifier reference point must hence be established via adequate filtering capacitors.

The corresponding situation is shown in Figure 9.2(c) for the less intuitive case of an amplifier with a pMOS input transistor. Note that, in this case, the critical voltage for the amplifier is the positive power supply and that the capacitor used to link the latter to the detector power supply must be capable of sustaining the high voltage needed to bias the sensor. While it is the job of the system designer to ensure that the detector and chip power supplies are clean enough, the integrated circuit designer should make any effort to prevent corruption of the amplifier ground or supply internally by noisy circuits put on the same chip. Techniques to prevent such problems are presented in Section 9.4.1.

### 9.2.2 Sensor Dark Current

In many semiconductor radiation detectors, the sensitive element is a simple diode that is reverse biased. Therefore, a DC current due to the thermal generation of free carriers is present and, in the case of direct coupling* between the sensor and the front-end electronics, flows into the input stage. Usually this current is very small (<1 nA) and can be accommodated by the input stage. In a few cases, however, the leakage current may increase dramatically. Two common examples are when the sensor operates in a harsh radiation environment that damages the bulk material (like in silicon detectors employed in high-energy physics experiments) or when the interesting signal is superimposed on an intense background (think of an infrared diode used in a TV remote-control set that is exposed to the ambient light). In such situations, the leakage current goes up easily by two to three orders of magnitude and, if left uncompensated, shifts the DC bias points in the front-end amplifier, thereby

* In some cases, in particular with silicon microstrip detectors, decoupling capacitors are fabricated directly on the sensor.

compromising its performance. A solution to minimize this undesired effect must then be adopted.

A simple possibility is to insert a decoupling capacitor between the amplifier stages, but integrated capacitors with an adequate linearity generally require a large silicon area. A more elaborate option is shown in Figure 9.1. We notice here that an additional feedback loop, consisting of a transconductance amplifier, is implemented around the second stage. The transconductor regulates its output current in such a way that the difference between the output node (VBL in Figure 9.1) and the reference voltage $V_{ref}$ is minimized. The residual error depends on the loop gain. In the ideal case, the leakage compensation loop should act only on DC or very slow components to avoid affecting the signals of interest. Therefore, the bandwidth of the differential amplifier should be small, in the range of 1 kHz or less. Such a low bandwidth is obtained by reducing the bias current of the transconductor to the nanoampere level and dumping it with a capacitor. In the case of a first-order loop, a detailed small-signal analysis reveals that the transfer function is equivalent to the one obtained with a simple AC coupling. However, this alternative approach has three advantages:

1. The low cutoff frequency depends on the transconductance of the differential amplifier and on the value of the filtering capacitor. Very low cutoff frequencies can hence be obtained by reducing the current in the differential amplifier while using a capacitor of moderate size.
2. The DC level of the output can be locked to the reference voltage and can easily be set to the most appropriate point for a given application, increasing flexibility of operation.
3. A nonlinear element (usually a buffer with very small slew rate) can be added in the loop, so that fast pulses are clipped before being presented to the transconductor. This reduces the shift of the baseline at high rates, otherwise unavoidable in a purely linear, AC-coupled system.

Called a *baseline holder*, this circuit entails interesting design issues that go well beyond the scope of this chapter. An exhaustive discussion can be found in De Geronimo et al. (2000).

The careful reader may have noticed that, with the arrangement shown in Figure 9.2, the first stage of the amplifier is still DC-coupled, so the sensor leakage current, flowing into the feedback resistor $R_f$, will shift the DC level at the preamplifier output. However, if the gain in the first stage is small enough, the leakage current can be absorbed without impairing significantly the dynamic range. In such a case, a compensation loop around the second stage will suffice, with the additional benefit of obviating the need for extra transistors that, if connected to the input of the first stage, may have a greater impact on the noise.

Up to now, we have considered dark currents that are interpreted by the front end as continuous or slowly variable components that can be well separated from the signal to be amplified. For some sensors, like standard photomultipliers or their solid-state counterparts, the silicon photomultipliers, the dark-current specification has to be looked at more carefully. In these sensors, in fact, a thermally emitted electron

triggers the avalanche multiplication, resulting in a fake signal similar to the one determined by a genuine photoelectron. The dark current is then the average current due to the superposition of many of these pulses. If the front end is fast enough and designed to be sensitive to a single photoelectron, the "dark pulses" will be treated as normal signals, creating an undesirable background. The rate of the dark pulses, rather than the resulting average current, becomes then the important figure to be considered in the design.

### 9.2.3 Noise

The noise for an integrated charge-measuring system is specified in terms of equivalent-noise charge (ENC). The required ENC varies widely among different applications. Fewer than 10 electrons rms may be needed in detectors for X-ray astronomy, while 2000 electrons are acceptable in fast readout circuits for photomultiplier tubes. The issue of minimizing the impact of the unavoidable sources present in a front-end chain, such as transistor thermal and flicker noise and detector leakage current, has been deeply addressed in the literature (see, for instance, Spieler [2006] and references therein). However, a complex charge-measuring ASIC incorporates circuits other than the analog front end, including digital gates switching between the power-supply rails. Furthermore, hundreds to thousands of these chips are usually employed to read out a large detector. The challenge is then to preserve the noise of the analog front end from being affected by other blocks hosted on the same chip or from being degraded when the chip that works well on the test bench, is incorporated into the final system. These issues will be the topic of Section 9.4.

### 9.2.4 Dynamic Range and Linearity

These specifications are sometimes sources of misunderstanding between the system builder and the ASIC designer, so we try to clarify them with the help of a practical example. We refer again to Figure 9.1. Suppose that the equivalent input noise specified for the circuit is 500 electrons rms and the gain is 50 mV/fC. The chip is powered with a single-rail 2.5-V supply, while the output stage of the analog front end can swing between 0.5 and 2 V. The output noise will be 4 mV rms and the dynamic range, defined as the ratio of the maximum signal to the noise, will be 375, or 51.5 dB. This is between an 8-bit and a 9-bit ADC, so let us examine the implication of either choice. If we select a 9-bit ADC, the LSB (least significant bit) will be 3 mV. In the optimistic assumption that there are no other noise sources in the system, the quantization noise, which is given by $\mathrm{LSB}/\sqrt{12}$, will contribute to the overall noise by only 2%. If we choose an 8-bit ADC, the quantization noise will be 1.7 mV rms, and it will give an 8% contribution to the final noise. However, if the converter has to be put on chip, it might be worth accepting the penalty, since the complexity of an ADC tends to grow exponentially with the number of bits.

High-resolution ADCs are off-the-shelf components at low prices and are often included in readout systems for detectors. When one moves to a more integrated system in which the converter in embedded on the front-end chip, it is natural to ask to maintain the same resolution, and the difficulty involved in the development of a

high-performance ADC is often underestimated. In defining the specifications of a fully integrated chip, the impact of the ADC resolution on the final system performance should therefore be assessed with careful system simulations, as a simplistic choice of the ADC may lead to investing precious design resources in a noncritical direction.

In a front-end chip, linearity is usually measured by feeding the circuit with input test pulses of different amplitudes and acquiring the corresponding amplitude of the output. The nonlinearity is derived from the deviation of the experimental points from the best-fit line. The performance of an ADC is quantified, among other parameters, by its differential nonlinearity (DNL) and integral nonlinearity (INL). An INL of less than 1 LSB for an 8-bit ADC means that the maximum nonlinearity of the converter will be 0.4% of the full-scale range. In most applications, such a figure is adequate, since imperfections at this level will be masked by statistical fluctuations intrinsic to the signal generation process in the sensor.

Some applications require the front-end electronics to have a dynamic range as high as 15 bits, but, for the reason just mentioned, the nonlinearity can be coarser. A good example is calorimetry in high-energy physics experiments. In this case, an ADC with lower resolution can be coupled to a multigain front-end amplifier, adapting the amplitude of the incoming signal to the full scale of the converter. The resulting multirange system is generally easier to design than a fully linear one incorporating an extremely high-performance ADC. An interesting example of this approach, reported in Raymond et al. (2005), will be discussed in Section 9.3. It is important not to translate automatically a specification on dynamic range into one on linearity, but it is necessary to understand separately the relevance of both for the system under consideration.

### 9.2.5 Event Rate and Efficiency

One of the most important requirements in a front-end chip is the rate of the pulses to be processed, since it determines the speed and has a direct effect on the power consumption of the circuit. In general, the arrival time of the input pulses is not uniform. This fact leads to the use of multibuffering schemes to relax the requirement on the dead time allowed to the single elements while minimizing data losses. As seen in Figure 9.1, the digitized pulses are stored in an output memory. The memory is organized in four slices, each one capable of accommodating the data corresponding to a full event. Let us suppose that the slices are periodically read out, that the time to read one slice is 2 μs, and that the incoming events follow a Poisson distribution with an average rate of 250 kHz. We can easily calculate how many events would be lost if only one slice in the memory were implemented. The probability of observing $n$-times an event with average value μ is in fact given by

$$P(n) = \frac{\mu^n e^{-\mu}}{n!}$$

In our example, $\mu = 0.5$, and the chance of not losing an event is given by the probability of having zero pulses during the time the memory is busy, i.e., 60%. The situation improves if more memory slices are added. In Figure 9.1, four elements are used, so that if an event arrives within the 2 μs needed to empty the first memory slice, it can be stored in another unit. The probability of losing one event becomes now the probability of having five consecutive pulses within 2 μs, which drops to 0.3%. Especially in high-rate applications, multibuffering is an important feature, because it allows us to "derandomize" the arrival time of the input pulses and to tune the following circuitry on the average event rate rather than on the peak one. This leads to design simplifications that are particularly important when the circuit has to be implemented as an ASIC. Data loss leads to a reduction in statistics and, hence, to longer measurement times. This is particularly detrimental in some applications, like nuclear and high-energy physics experiments, where rare events are looked at, or in nuclear medicine, where longer measurement times are associated with higher radiation doses delivered to the patient.

### 9.2.6 Power Consumption

Power consumption is usually specified in milliwatts per channel. Power dissipation—traditionally a big concern in vertexing and tracking systems employed in colliding-beam experiments—is now also becoming a relevant factor in other areas. A good example is provided by PET scanners based on solid-state sensors like avalanche photodiodes, which are quite sensitive to temperature variations. Hence, the heat produced by the front-end electronics must be efficiently removed, and in cases where the front-end chip is too power hungry, the cooling system becomes cumbersome. Power consumption may vary according to the particular operation the ASIC is performing at a given time. Therefore, it should always be specified, distinguishing between average power consumption, which is relevant for the design of the cooling system, and peak power consumption, which is the factor to be considered for the sizing of the electrical components (printed citcuit board [PCB] traces, filtering capacitors, power supplies, etc.). Particular care must be paid to the design of an adequate power-supply network inside the ASIC itself, since metal lines used on chip are relatively resistive. As a reference, the resistance of a trace measuring 1-mm long and 10-μm wide is on the order of 5 Ω, so nonnegligible voltage drops may manifest along the trace if a current beyond a few milliamperes flows through it.

## 9.3 ARCHITECTURES FOR INTEGRATED CHARGE-MEASURING SYSTEMS

### 9.3.1 Full-Flash Architectures

In full-flash architectures, such as the one reported in Figure 9.1, the digitization is introduced immediately after the front-end amplifier. With this approach, the information is transferred as early as possible in the digital domain, which offers more robustness and the possibility of applying early online powerful digital signal-processing techniques.

An example of such an architecture can be found in Yeom et al. (2007), which describes a nine-channel front-end ASIC developed for APD readout in a positron emission tomography (PET) application. Each channel of the chip consists of a transimpedance amplifier followed by a second stage with variable gain. This feature is particularly desirable in systems involving pixelated photodetectors, such as APDs or multianode photomultipliers, because the gain spread among pixels of the same sensor can be as high as 30%. The variable gain in the amplifier is then exploited to equalize the response of different channels. The output of the analog stage is immediately digitized at 50 MHz by a 6-bit ADC built with a folding topology (Razavi 1995; van Valburg and van de Plassche 1992), which allows the design of an $n$-bit full-flash converter with less than the $2^n - 1$ comparators needed in the traditional scheme. Produced in a 0.35-μm CMOS process, the ASIC dissipates 200 mW per channel.

An ADC requires a significant amount of digital circuitry. In full-flash architectures, the converter and its logic are active while the input amplifier is processing the signal. As we will see in Section 9.4, a monolithic implementation allows the switching noise generated by the ADC to be easily injected into the front-end stage, thereby corrupting system performance. This risk is reduced for full-flash architectures that are implemented in the form of "chip sets," with the amplifier and the ADC fabricated on two separate dies. The front-end electronics of the electromagnetic calorimeter (ECAL) of the compact muon solenoid (CMS) experiment provides an interesting example of this concept (Raymond et al. 2005).

This electronics architecture has been designed to serve two types of sensors: avalanche photodiodes (APD), employed in the barrel part where the magnetic field is stronger, and vacuum phototriodes, which instrument the end caps on which the radiation load is more severe. The APDs present a capacitance of 200 pF to the front end, while for the phototriodes the capacitance is 50 pF, mostly dominated by stray contributions. Depicted in Figure 9.3, a single channel is formed by a chip with a multirange amplifier, with the multigain preamplifier (MGPA) followed by a second ASIC hosting a multichannel ADC. Both circuits are fabricated in the same 0.25-μm CMOS process. The MGPA is made by a low-noise preamplifier coupled to three shaping stages, each one having a different gain (1×, 2×, 12× with respect to the preamplifier gain). The output of each shaper is fed to a dedicated ADC

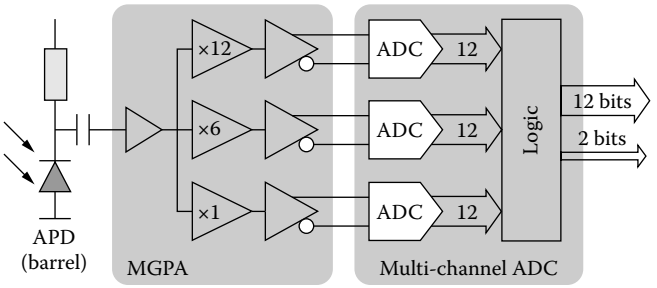



**FIGURE 9.3** Block scheme of the front-end electronics for the CMS ECAL calorimeter, with the front-end chip and the fast ADC residing on two separate ASICs. (From Raymond et al. [2005, p. 756]. With permission.)

channel on the converter chip, which is based on an industrial IP (Minderico et al. 2003). The ADC has a resolution of 12 bits, which is obtained by cascading a flash stage with 2.5-bit resolution and a 10-bit pipeline ADC with 1.5-bit/stage topology that digitizes the residue of the flash. The ADC is followed by a digital circuit that selects the code with the highest nonsaturated value and sends it to the serial output port. Low-voltage differential signaling (LVDS) is used for the digital I/Os to minimize switching noise. The multigain capability of the MGPA is exploited to expand the dynamic range of the system, which has a 12-bit resolution, to 16 bits, and the selection of the suitable gain range to be used for further processing is done a posteriori by a simple digital circuit. This can be contrasted with other applications in which range selection is performed at the analog level (Anghinolfi et al. 1995; Bonvicini et al. 2007).

The flexibility of the digital postprocessing makes it easy to handle conditions that would be problematic for a pure analog technique. For instance, a hysteresis mode allows digitizing to continue in the same channel of the peak for the samples that occur in the tail after the peak itself, in order to avoid large discontinuities on the same signal. The power consumption of the ECAL frond-end electronics is 600 mW for the MGPA and about 500 mW for the ADC chip.

A very good demonstration of the potential offered by digital signal conditioning in a charge-measuring application is given by the front-end electronics of the Time Projection Chamber of the ALICE experiment at CERN, which studies ultra-relativistic heavy ion collisions at the LHC.

The chip set consists again of two ASICs, each one integrating 16 parallel channels. The first chip is a purely analog front end and is implemented in a 0.35-μm CMOS process. Each channel is made of a charge-sensitive amplifier, followed by a pole-zero cancellation network and a two-stage T-bridge filter to achieve a fourth-order shaping function. The second stage of the filter is also used to convert the single-ended signal coming from the previous stages to a fully differential one, suitable to drive the ADC. The differential outputs are buffered individually, so that the DC level of each one can be independently regulated.

The amplifier is designed for a total conversion gain of 12 mV/fC, a full width at half maximum (FWHM) of the output pulse of 190 ns, and a power consumption of less than 20 mW per channel. The output signal of the front-end amplifier is fed to the second ASIC. Called ALTRO (Alice Tpc ReadOut) (Bosch et al. 2003), this chip is fabricated in a 0.25-μm CMOS technology. It embodies in each of its 16 channels a 10-bit pipeline ADC followed by a multistage digital signal processor (Figure 9.4). The first stage of the chain is used to perform preliminary processing on the incoming data stream delivered by the ADC. Gain equalization, nonlinearity corrections, and offset subtraction are applied at this level. The offset is calculated by averaging baseline samples taken outside the acquisition window, which is defined by an external trigger signal. In this way, slow baseline drifts due, for instance, to temperature variations can be compensated for.

A dedicated memory stores a fixed pattern that can be subtracted to remove systematic disturbances superimposed on the input signal. In addition, the memory can be used to provide a known pattern to the next stages. This feature makes it possible to test the following part of the processor by injecting a digital "calibration signal,"

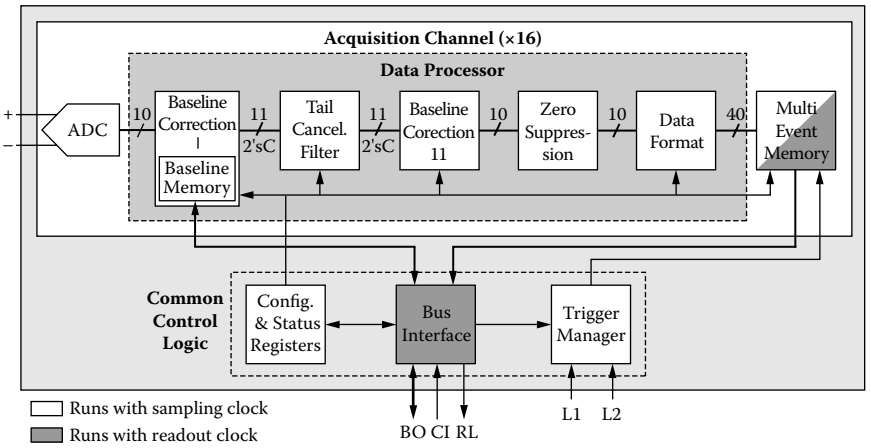



**FIGURE 9.4** Block scheme representing one processing channel of the ALTRO chip and the common control logic. A total of 16 identical channels are accommodated on the ASIC. (From Bosh et al. [2003, p. 2461]. With permission.)

without the need of stimulating the analog section. The preprocessed data are then fed to a digital filter (Mota et al. 2000). The function of this block is to minimize pileup of signals by cutting the long tail due to the slow ion motion that is typical of gaseous detectors. Implemented as a third-order infinite impulse response (IIR) filter obtained by cascading three first-order filters, this circuit is one of the most interesting parts of the chip. Traditionally, in fact, tail cancellation was obtained by pole-zero analog networks (Baur et al. 1998), whose performance is limited by the accuracy of the devices used to build them. In the digital domain, on the other hand, the precision of a calculation depends only on the length of the data words. Furthermore, the filter can be easily reprogrammed and adapted to different experimental scenarios by changing the values of a few coefficients. After the tail correction, the data are passed to another circuit that inspects the baseline to look for nonsystematic variations determined by pickup noise. Such components are calculated by mean of a moving-average filter and subtracted. At this point, the interesting pulses are nicely superimposed on a flat baseline, so zero-suppression can be reliably applied by means of a simple threshold. In order to avoid occasional noise glitches, at least two consecutive samples must be above threshold. A time stamp is added to allow proper event reconstruction. After being formatted in a 40-bit word, the final data are stored in a multihit memory that can accommodate up to eight consecutive events.

Two further aspects of this chip deserve attention. First, the circuit uses two different clocks: one for controlling the ADC and the signal-processing circuits and one for managing the readout part. In this way, it is possible to adapt independently the speed of the two sections of the ASIC to the experimental needs. When possible, one could use the ADC at lower speed to improve resolution and reduce power consumption while still running the output stage faster to maximize the data-transmission bandwidth. Additionally, data readout is stopped during the acquisition of a new

event to reduce the interference of the output stages on the ADC. This is particularly important, since the chip employs CMOS output pads with 2.5-V swing.

A second interesting point is that the sampling clock of the ADC is out-phased with respect to the clock that controls the rest of the logic. In this way, when the logic switches, the samples have already been acquired by the converter, and the risk of having them corrupted by some internal pickup of the digital activity is reduced. The power consumption of the chip, when the ADC is sampling at 10 MHz and the chip is read out at 60 MHz, is 20 mW/channel, while the ADC alone consumes 43 mW when it is used at its maximum speed of 40 Msamples/s. Further details on this ASIC can be found in Bosch et al. (2003).

### 9.3.2 Architectures Based on Analog Memories

A common pattern that emerges from the circuits discussed in the previous subsection is that full-flash architectures are characterized by a modest integration density coupled to a relatively high power consumption. The fast ADCs needed in these systems are, in fact, power-hungry components. As a rule of thumb, one can assume that a 10-bit converter needs between 0.5 and 1 mW per Msample, leading to a power dissipation between 20 and 40 mW for a 40-MHz ADC.

In many applications, the power demanded by an early digitization executed immediately after the front-end amplifier can not be afforded. One common example is given by the silicon tracker employed in high-energy physics experiments, where front-end chips with 64–128 channels and a power budget of 2–4 mW per channel are necessary. Another case is found in the front-end electronics of the highly integrated PET scanners envisaged for multimodality imaging and based on highly segmented photon detectors. In these applications, sampling and temporary analog storage, which require much simpler hardware, can be decoupled from digitization. The former have to be carried out, in fact, for all possible events, while the latter is needed only for those events that actually occur. In colliding-beam experiments, for instance, two beams of accelerated particles are smashed against each other at a given frequency (25 ns in case of the LHC), which hence determines the sampling rate of the associated detector systems. However, the frequency of the truly interesting events created in the collisions, and for which digitization is indispensable, is much smaller (100 kHz or less). Similarly, in a PET detector, a fast front end is desirable to preserve good timing information, but the event rate is normally limited to a few kilohertz per channel. Hence, fast sampling and analog storage can be coupled to a slower ADC per channel, or, alternatively, to a fast ADC shared by many channels. The converter can also be put outside the front-end chip and located in areas in which power dissipation is less critical.

Analog values can be sampled and held with compact cells consisting of capacitors and switches. Many of these cells can be put in one channel to keep the information stored until an external trigger signal initiates the readout and to derandomize the arrival time of the events. In this way, the readout can take place at the average rate without incurring a significant data loss. Called *analog memory*, this class of circuits has found widespread use in the design of integrated circuits for radiation detectors. ASICs with up to 512 storage cells per channel have been described in the

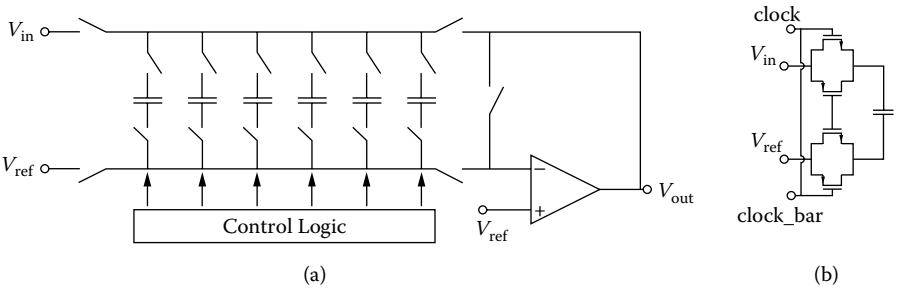



**FIGURE 9.5** (a) Simplified schematic of an analog memory channel employing a voltage-write, voltage-read topology; (b) detail of a sampling cell.

literature (Baur et al. 1998). Figure 9.5 presents a typical analog memory channel and the details of one storage cell. Despite its simplicity, the design of these circuits entails a few relevant issues.

A key point is the choice of the sampling capacitor. Many technologies with a feature size of 0.35 μm or coarser offer two layers of polysilicon to form very linear capacitors with a density of 1–2 fF/μm$^2$. Hence, a 500-fF capacitor, which is a typical value for a storage cell, requires an area of $20 \times 20$ μm. In processes of the quarter-micron generation and beyond, linear capacitors are instead obtained by two dedicated layers of metals. Yield considerations generally prevent the use of these components in large analog memories. The reason is as follows.

Modern CMOS technologies constrain each layer of material used in chip fabrication (diffusions, polysilicon, and metals) to cover a given chip area. For instance, a typical requirement is that any metal layer present in the process must occupy between 25% and 70% of the silicon area, with a distribution as homogeneous as possible. Violations of these rules result in poor planarity of the oxides separating the different layers, with possible failures of the vias that have to connect them in selected points. A long chain of cells, replicating over many equal channels, is likely to violate such a rule. Even if a small prototype works successfully, the yield can be very poor when a full-size chip is produced. Furthermore, special processing steps are often involved in the fabrication of metal–metal capacitors, and thus their use may be subject to quite severe restrictions. One is then forced to employ MOS capacitors, which in deep-submicron processes offer a very high density (5 fF/μm$^2$ being a common figure), but also exhibit a significant voltage dependence. Therefore, the impact of the capacitance variation with the applied voltage must be minimized to avoid exploiting the MOS device to perform current-to-voltage conversions. For example, if a voltage is stored in the capacitor by driving the cell with a low-output impedance amplifier, than the cell *should not* be read out by directly discharging it with a constant current source and counting the time taken by this operation. This would, in fact, measure the charge stored, which, due to the voltage dependence of the MOS capacitor, will be related in a nonlinear way to the written voltage.

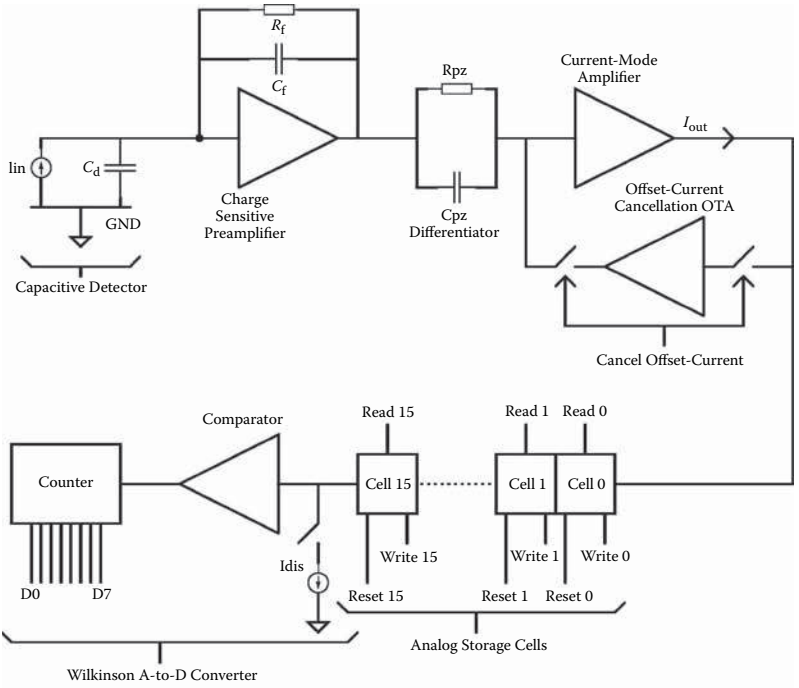
Another potential problem arises from the charge injection of the switches. The charge stored in the conductive channel of a transistor is proportional to $V_{GS} - V_{TH}$, where $V_{GS}$ is the gate-source voltage and $V_{TH}$ is the threshold voltage. Since it has

its source grounded, the switch put in the signal return path will inject a constant charge. The switches in the signal path, on the other hand, will store in their channel a charge depending on the sampled voltage. The problem is further exacerbated by the modulation of the threshold voltage due to the body effect. To maximize the linearity, the switch in the signal path can be removed, as discussed in Haller and Wooley (1994). However, this affects the useful dynamic range. In fact, in this case, the top plate will always be connected to the front-end amplifier, and its voltage will also change when the cell is not written. In order to preserve the charge, the bottom plate of the capacitors, which are in the storage mode, will change accordingly, and it may be driven below ground, putting the associated switches back into conduction. In practice, good results in terms of dynamic range and linearity have been obtained by maximizing the symmetry of the circuit with switches connected to both plates (Anelli et al. 2001). Attention must also be paid to the design of the circuitry that generates the digital signals that drive the switches. The goal here is to avoid charge sharing between two consecutive cells and to minimize digital feed-through. The reader is referred to Anelli et al. (2001) for a more in-depth treatment on the implementation of analog memory in submicron CMOS technologies.

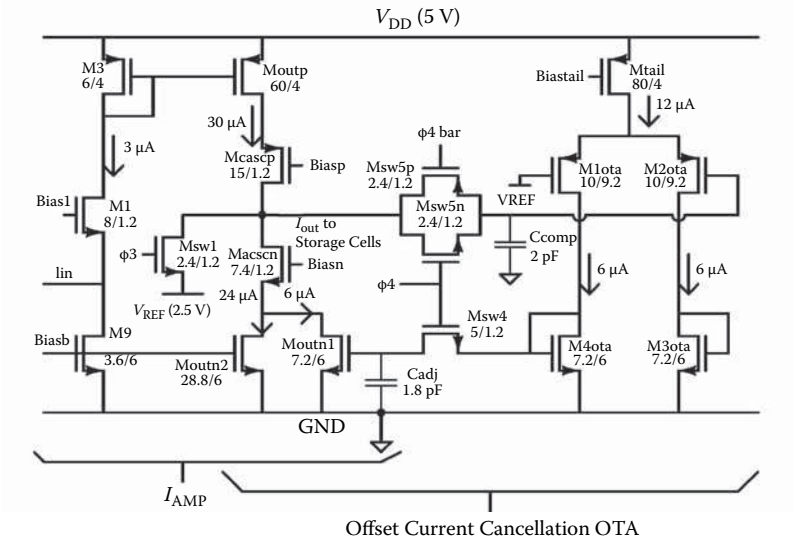
To illustrate further the use of analog memory in charge-measuring integrated circuits, we now discuss a few practical cases. The circuit reported in Tedja et al. (1995) provides an instructive example of an integrated front end employing this concept. Implemented in a 1.2-μm CMOS process, the chip was designed to read out silicon-strip detectors with 30-pF sensor capacitance and optimized for a particle-physics experiment in which collisions took place every 144 ns. Figure 9.6(a) shows the block diagram of one channel. The input stage consists of a charge-sensitive amplifier followed by a pole-zero cancellation network. Here, the sensor current is amplified by the ratio $C_{pz}/C_f$ and shipped to the second stage. The latter is a current amplifier that combines a common-gate topology with a current mirror and has a gain of 10. The output branch of the current mirror, which delivers the signal current to the analog memory, is built with a telescopic cascode architecture to increase its output impedance. Another amplifier is inserted in the feedback path of the second stage. The purpose of this additional element is to suppress the DC offsets and to set the high-impedance output of the second stage to a stable bias point. The technique can be understood with the help of Figure 9.6(b) and works as follows.

During the calibration phase, the switches are closed. The feedback amplifier has a differential architecture and, combined with the output branch of the second stage, forms an operational transconductance amplifier (OTA), connected as a unity-gain buffer. The OTA locks the output node to a fixed reference voltage. During calibration, the front end is not able to process an input signal. When the switches are opened, the differential amplifier is removed from the feedback path, and the front end is put back into the acquisition mode. A voltage is stored in a dedicated capacitor ($C_{adj}$ in Figure 9.6b) and converted by a transistor ($M_{outn1}$) to a current that is added to the output branch of the second stage to preserve the same DC conditions established during calibration.

The output of the front-end amplifier is first shorted to a reference voltage to discharge its associated parasitic capacitance and then integrated for 128 ns in a storage capacitor in the analog memory. Performed in a time interval of 16 ns, the former operation is necessary to prevent signals that were integrated on a given time window




(a)




(b)

**FIGURE 9.6** (a) Front-end channel with a current-mode amplifier followed by an analog memory and a Wilkinson ADC. (b) Detailed schematic of the second-stage current-mode amplifier with the OTA for dynamic offset compensation. (Adapted from Tedja et al. [1995]. With permission.)

from being partially summed in the next one. Samples are therefore acquired every 144 ns. Each integration cycle is performed on a different sampling cell to avoid overwriting a potentially interesting event. The analog memory consists of 16 cells, so the information can be stored for 2.3 μs. If, during this time, the sample is validated by an external trigger command, the interesting cell is put in readout mode, and its content cannot be overwritten. The other capacitors may still be used for acquiring new events. The absence of a trigger within 2.3 μs determines the overwriting of the cell by a new sample. The total number of cells chosen for the memory is hence defined by the latency of the trigger signal, with the circuit acting as an analog delay line. When a cell is tagged for readout, the charge stored in the capacitor is digitized with a Wilkinson ADC: A constant current source discharges the capacitor, and the time needed for this operation is measured by an 8-bit counter driven by a 62.5-MHz clock. Hence, the conversion of a full-scale signal lasts 4 μs. The full channel works as a charge digitizer: Only one sample per event is taken, and the information on the shaper of the waveform at the preamplifier output is not preserved.

It is important to observe here that, when digitization takes place, some extra digital logic is working on the chip, and its switching could induce noise on the analog section. A possible alternative is avoiding concurrent digitization and sampling, with the penalty of introducing a dead time into the system. Circuits that are capable of performing simultaneously the digitization and readout of stored samples together with the acquisition of new ones are usually referred to as "dead-time free."

The reader should note that all the information is processed by the ASIC in the current domain, without requiring precise current-to-voltage conversions. This has two key advantages. First, the gain of the system depends on the ratio between components of the same type. The matching between the parameters of two integrated devices of the same nature is much better controlled than their absolute values, so the gain of the channel can be quite precise and does not need to be calibrated. Additionally, one stores in the capacitors and reads back from them the same physical quantity, which in this particular case is a charge. Thus, as noted previously, the linearity of the capacitors is not a major issue, and they can be implemented using compact but relatively nonlinear MOS transistors. The chip dissipates 1 mW of power per channel, most of which (620 μW) is invested in the charge-sensitive amplifier to keep the input referred noise below 1800 electrons rms. The circuitry needed for the Wilkinson ADC dissipates only 150 μW. It is interesting to compare this figure to the power dissipated by a full-flash architecture working at the same speed. An ADC with a sampling rate of 7 MHz typically takes between 3.5 and 7 mW. Including the front-end amplifier, the total power consumption would be between 4 and 7.5 mW per channel, the extra power being mostly spent in the digitization of samples that are soon discarded.

We now turn our attention to another example of a charge-measuring ASIC based on analog memory. The circuit was developed to read out the silicon drift detectors (SDDs) of the aforementioned ALICE program at CERN. A first 32-channel prototype of this chip is described by Mazza et al. (2004), while some details on the final 64-channel version installed in ALICE can be found in Rivetti et al. (2005).

SDDs (Gatti and Rehak 1984) exploit segmentation of the charge-collecting electrode in one direction and measurement of the charge drift time in the other to obtain

an unambiguous two-dimensional position measurement with a reduced number of channels. In ALICE, a single SDD sensor has an active area of $7 \times 7.5$ cm and is read out on both sides by an array of 256 anodes with 294-μm pitch (Rashevski et al. 2002). Analog readout is exploited to calculate the center of gravity of the charge cloud in both the anode and the drift (time) directions. This leads to a resolution of 30–40 μm in both coordinates (Nouais et al. 2003). To reconstruct the center of gravity in the time direction, the information on the signal shape is preserved by taking multiple samples on the same waveform. A simplified schematic of the chip showing the relevant building blocks is reported in Figure 9.7.

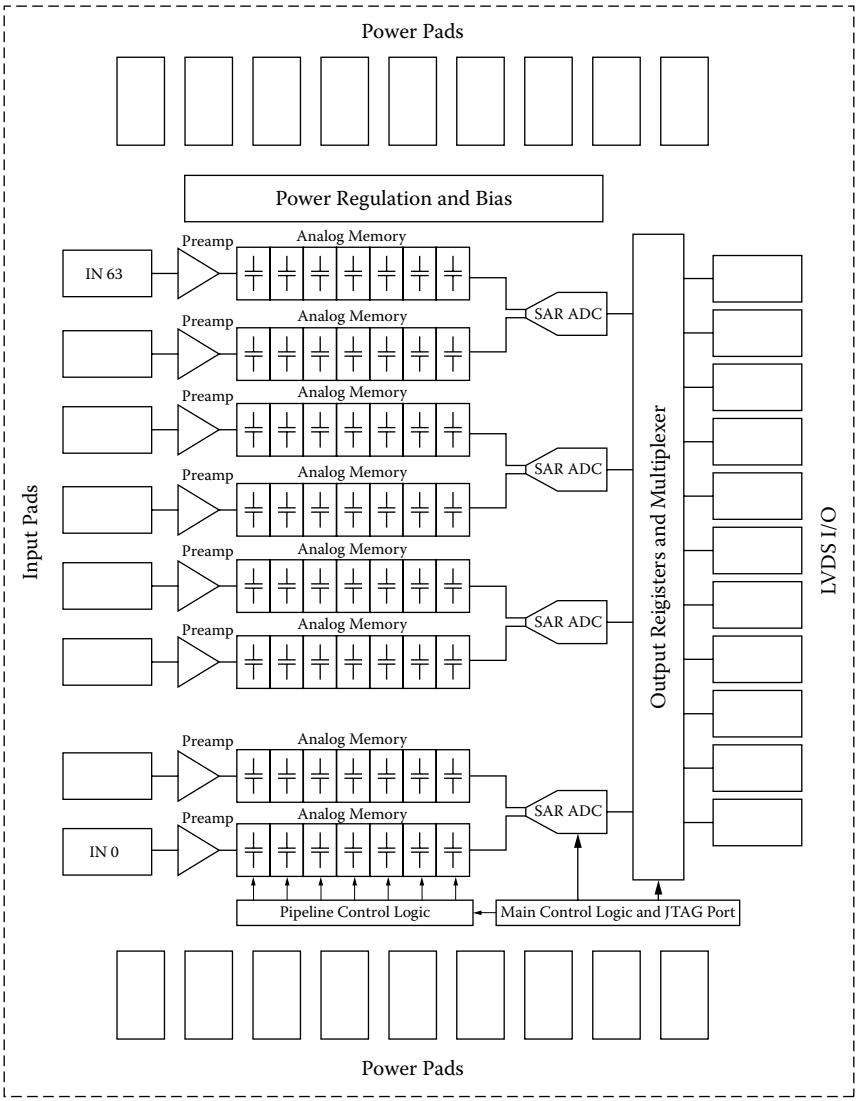



**FIGURE 9.7** Block diagram of the front-end ASIC developed for the ALICE SDD.

In each channel, the signal coming from the sensor is amplified by an analog front end with 40-ns peaking time. The front end converts the sensor current into a voltage with a conversion gain of 50 mV/fC. The resulting output is sampled by an analog memory with 256 cells that is driven by a 40-MHz clock. The number of cells needed is defined by the maximum drift time foreseen (6.4 μs) divided by the sampling time (25 ns). In other words, the depth of the memory must allow the chip to store a full picture of the event that has occurred in the sensor. The ASIC relies on an external trigger to select valid data, and trigger latencies up to 6.4 μs can be accommodated. The trigger starts the digitization phase. In this mode of operation, the cells are sequentially put in feedback to an operational transconductance amplifier, and the voltage stored into them is read back. Once more, the memory is not used to perform any conversion that would rely on the linearity of the sampling capacitors, which are hence implemented with compact MOS devices.

The OTA drives the analog-to-digital converter. To prevent the coupling of the ADC switching noise into the front end, digitization and sampling are mutually exclusive. This also simplifies the architecture of the analog memory, at the expense of introducing a dead time. To minimize the dead time, a successive-approximation ADC, which is significantly faster than a Wilkinson, is used. The converter, described by Rivetti et al. (2001), has a resolution of 10 bits and operates with a full scale of 1.4 V, so the LSB is 1.4 mV. Clocked at 40 MHz, the ADC needs 250 ns to complete a conversion. To this time, one has to add the settling time of 370 ns that is needed by the read amplifier. The conversion of one sample therefore lasts 620 ns, to be compared to the 25.6 μs that would be required by a Wilkinson ADC operating in the same conditions. One ADC is shared between two adjacent channels, so the time to read the full memory is 317 μs.

The ASIC also has a "fast mode," in which the sampling clock of the analog memory is halved. This allows cutting the number of samples and hence the dead time by a factor of 2, at the expense of a somewhat lower but still acceptable resolution. The final 64-channel chip embodies 32 ADCs. Working in parallel, the converters present a severe load to the two reference voltages, which are hence provided by two low-drop-out regulators (LDOs) incorporated directly on chip. The digitized data are sent off the chip via a 20-bit parallel bus working at 40 MHz. Because conversion and data transmission can be concurrent, all the digital I/Os are differential. However, the power consumption of standard LVDS links was not affordable in this application. Taking advantage of the fact that, in the final system, data are passed to another ASIC located less than 1 cm away, the digital transmitters are implemented with low-current drivers that steer a current of 150 μA in a loop closed on a 2-kΩ resistor. The clock is the only signal received with a standard LVDS link. Implemented in a 0.25-μm CMOS technology, the front-end ASIC dissipates, per channel, 3.5 mW of power during sampling and 4.8 mW during digitization. This must again be compared to the 20–40 mW per channel that would be needed by a full-flash approach with the same sampling speed.

As our last example of a circuit employing analog memory, we discuss briefly a recently reported chip developed for an APD readout in a positron emission mammography application (Rodrigues et al. 2009). The need of reading more than 12,000 sensing elements has motivated the development of this ASIC, which

incorporates 192 channels that process the signals delivered by six different APD arrays. In each channel, a front-end amplifier converts the detector current into a voltage pulse with a gain of about 20 mV/fC and a peaking time of 20 ns. The output of the amplifier is continuously sampled by an analog memory that has 10 cells, each implemented with a poly-poly capacitor. A fast comparator monitors the amplifier output and fires in case a signal above threshold is detected. The samples stored in the corresponding channel are then sent off the chip using one of the two available output buffers.

In parallel with the analog samples, the address of the firing channels is sent via a dedicated LVDS link. Two channels can be served simultaneously. If a third hit is detected while the output buffers are occupied, an error flag is generated. The samples are written and read back from the memory at the same frequency (50 MHz). Thus, the chip acts as a fast, self-triggered waveform sampler and multiplexer that reduces the number of lines from 192 to 8. In fact, four lines are needed for the differential transmission of the analog data, while the other four are taken by the two LVDS links that send the channel address. Interestingly, the uniformity of the comparator is so good that the same threshold voltage can be applied to all of them with an external DAC, without requiring fine tuning on chip. The fact that sampling and readout in the ASIC are concurrent does not introduce undesirable effects, and the noise of the system (1300 electrons rms) is defined by the intrinsic thermal noise of the front-end amplifier. The chip is fabricated in a 0.35-μm CMOS process and dissipates 3.5 mW per channel from a 3.3-V power supply.

### 9.3.3 Peak Detectors and Hold

An interesting alternative to waveform sampling is provided by peak detection and hold. In this scheme, the front-end amplifier is followed by a unidirectional circuit that tracks the amplifier output until the maximum is reached and stores the corresponding value in a capacitor. Called an *analog pulse stretcher*, this component is commonly available in the form of a stand-alone module for crate-based nuclear instrumentation. Its implementation in an integrated form in CMOS technologies is a not a trivial task. A detailed analysis of the design issues involved in this concept is beyond the scope of this chapter, but the interested reader can find an exhaustive treatment in the work of De Geronimo et al. (2002a).

In a related paper, the same authors (De Geronimo et al. 2002b) propose a novel scheme that implements an accurate peak detect-and-hold function in CMOS. Shown in Figure 9.8, the circuit is based on the concept of using the same components in two separate configurations, one optimized for peak detection and the other optimized for storage. In the peak-detection mode, switches S3, S4, S5, and S6 are open, while S1 and S2 are closed. The S6 switch allows setting the value of the voltage $V_h$ to a known baseline value. Waveforms in the key nodes of the circuit are shown in the right part of Figure 9.8. We notice that when the input signal is rising, an error voltage $V_{in} - V_h$ is sensed by the OTA, which tries to minimize this error. As a consequence, the voltage on $V_g$ is driven abruptly down, and a current flows in Mg and M1. The current in M1 charges the hold capacitor, whose voltage follows the input one. After the peak, the input signal starts its return to the baseline, so to follow it at $V_h$,

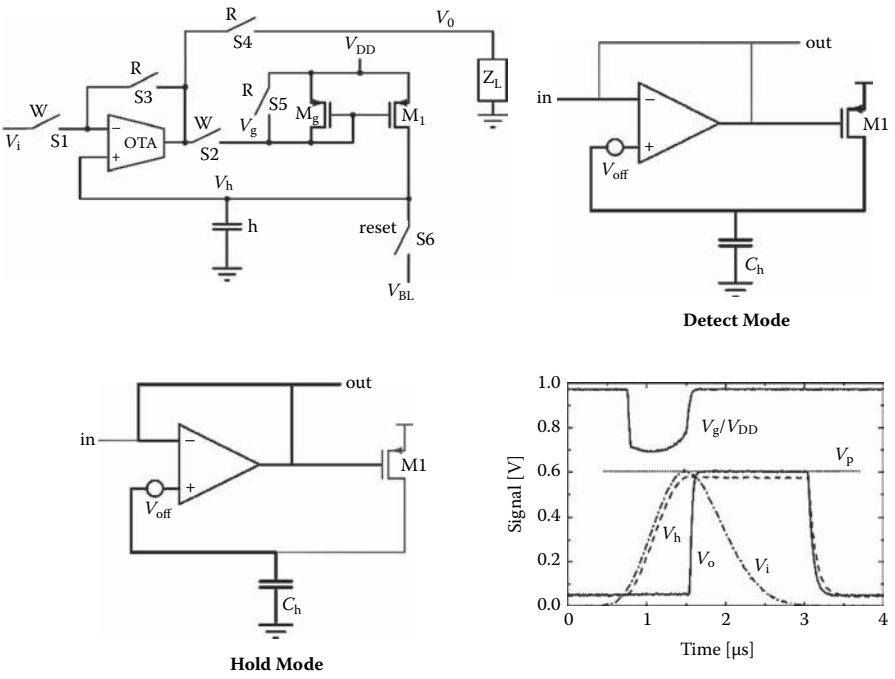



**FIGURE 9.8** Schematic of the two-phase peak detector and hold circuit (left). Details of the detect-and-hold mode configurations (center) and relevant waveforms (right). (Adapted from De Geronimo [2002b]. With permission.)

one would need to discharge the capacitor. However, the pMOS transistors can only source current, so the node $V_g$ is driven sharply up. Mg and M1 are quickly driven in the off-state, and the value of the peak is stored. A comparator is used to sense the second transition on $V_g$ and to change the configuration of the switches. Switches S1 and S2 are now opened, while S3, S4, and S5 are closed. The OTA is now put in a unity-gain configuration, with the buffer hold capacitor isolating it from the subsequent circuitry, thereby avoiding any loading effect that may corrupt the accuracy of the peak. A key point is that, since the same OTA is used in both phases, its offset is naturally subtracted, becoming inessential. Furthermore, since the second $V_g$ transition occurs on the peak, it is ideally independent of the signal amplitude and thus provides accurate timing information to time-stamp the event. The circuit is hence self-triggered.

Several cells can be used in one channel to provide multibuffering capability. To maximize the dynamic range, the OTA employs a rail-to-rail input stage. A linear dynamic range of 2.7 V is achieved while the circuit, which is implemented in a 0.35-μm process, is powered at 3.3 V and dissipates 3.3 mW. An absolute accuracy on the peak measurement of 0.2% was demonstrated using the circuit in a high-rate spectroscopy system based on CZT (cadmium zinc telluride) detectors (De Geronimo et al. 2002c).

### 9.3.4 Full-Flash versus Analog Derandomization: Some Remarks

It is apparent from the practical cases discussed in the previous subsections that circuits employing analog sampling and derandomization before the ADC offer significant savings in power consumption with respect to full-flash implementations. We see here an example of a more general fact. In acquisition systems, architectures based on slower processors with buffers in front are normally cheaper than their counterparts built with faster processors, since memory demands less complexity than processing. Furthermore, for a given number of bits, the difficulties encountered in ADC design tend to grow exponentially with the conversion frequency required. For a very demanding application, it can be more efficient to incorporate in the front-end chip a converter provided by a specialized design house rather than embarking on a custom development, which can be very time consuming.

While architectures based on analog memory or peak detection are generally more power efficient, the full-flash approach may have a clear advantage in self-triggering applications. Front-end circuits for radiation detectors are very often designed using single-ended circuit topologies. This is motivated by the fact that the sensor is single-ended, so even if a fully differential design style is employed inside the chip, the symmetry will break at the input stage. This has led many authors to prefer simpler single-ended architectures, which also offer some power savings. Thus, these circuits have limited common-mode rejection. A self-triggering architecture based on an analog memory requires a comparator at the output of the amplifier, and if some common mode is present, the threshold needs to be increased, which affects the sensitivity. On the other hand, if the data are immediately digitized and processed in pipeline, common mode noise could be calculated and subtracted digitally before deciding whether the hit has to be further processed or discarded. In circuits that are not self-triggered, this aspect is less critical, since the decision about the acquisition is made by another system and, once they have been digitized, the data can be filtered to suppress common mode components.

## 9.4 PRACTICAL DESIGN ASPECTS IN INTEGRATED FRONT-END SYSTEMS

We now address practical aspects that are fundamental to the design of large front-end ASICs that perform well in real systems. The complexity of the topics makes it impossible to give an exhaustive treatment in the span of a few pages. Thus the aim of this section is to make the reader aware of the key problems and to provide adequate references for a more in-depth study.

### 9.4.1 Minimizing Digital Noise

It is apparent from the examples in the previous section that multichannel charge-measuring ASICs are, in general, mixed-mode circuits, in which many digital gates share the silicon with sensitive analog components that must detect signal charges below 1 fC. By switching between the power-supply rails, the digital logic can inject through the common substrate unwanted disturbances that can easily corrupt the

performance of the analog blocks. Called *substrate noise*, this phenomenon is of great importance for all mixed-mode integrated circuits, and a wide reach of literature on this topic has developed along the years (Su et al. 1993; Blalack and Wooley 1995; Aragonès and Rubio 1999; Peng and Lee 2004; Badaroglu et al. 2004; Lan et al. 2006). To illustrate the problem, we refer to Figure 9.9. Here, two MOS devices are shown, one belonging to a digital circuit and the other to an analog circuit.

In Figure 9.9(a) the transistors are fabricated on a lightly doped (typically 10 ohm·cm) epitaxial layer grown on a highly doped substrate (0.01 ohm·cm), while in Figure 9.9(b) the whole substrate is lightly doped. Epitaxial wafers are frequently found in processes of the 0.35–0. 25-μm generation, while lightly doped ones are commonly used in 0.13-μm technologies and beyond. When the inverter is switching, three main mechanisms can inject parasitic currents into the substrate. First, electrons accelerated by the electric field in the device may have sufficient energy to create electron-hole pairs (impact ionization), with the holes flowing into the substrate. Second, current can be injected via the capacitance associated with the reverse-biased junctions. Third, the fast-current transients, which happen when the gate changes its state, create bounce in the power-supply lines due to the nonzero inductance of the connections. The bounce on the positive power supply creates current into the substrate via the capacitive coupling occurring between the substrate

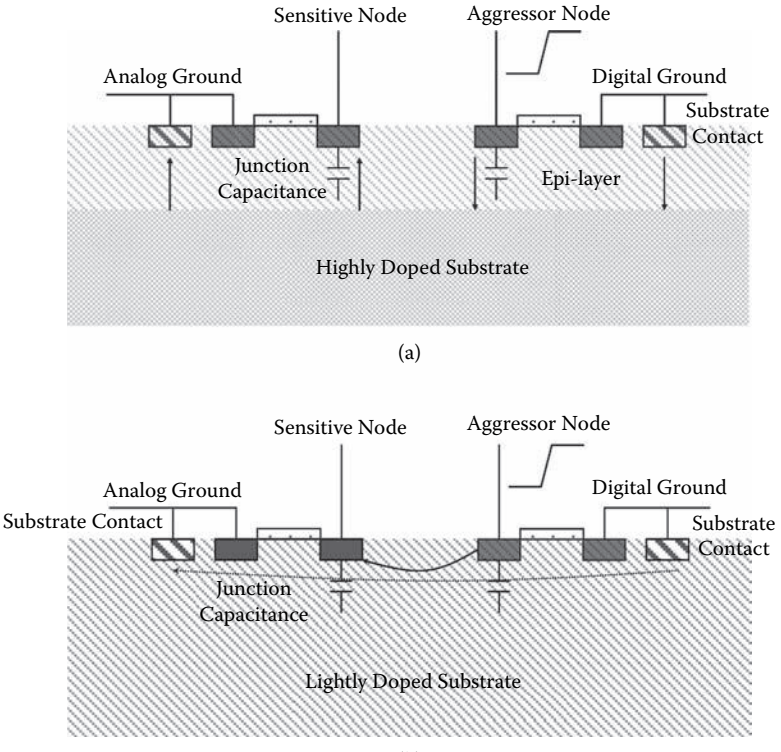



**FIGURE 9.9** Digital noise path in epitaxial (a) and lightly doped substrates (b).

and the *n*-wells in which pMOS transistors are fabricated, while the bounce on the ground couples to the substrate via the finite resistance of the substrate contacts. This last mechanism was found to be dominant in an experimental study conducted by Aragonès and Rubio (1999) and is confirmed to be the key player in substrate noise injection in deep-submicron technologies (Badaroglu et al. 2004). Parasitic currents flowing into the substrate can affect a critical device, like the input transistor of an analog front end, in three possible ways: tweaking the local bulk potential and hence the threshold voltage; coupling AC currents via the junction capacitances; and modulating the analog ground line, which is typically also used to bias the substrate in the analog portion of the chip.

The way that parasitic currents spread in a circuit heavily depends on the wafer type. In epi wafers, the currents quickly reach the highly doped part, which behaves as a lumped node. A disturbance in this node is easily picked up by the analog circuits through the substrate contacts. Therefore, in this case, protective guard rings around the sensitive circuits and physical separations between digital and analog blocks, although advisable, may not be very effective in reducing the induced noise. It has been suggested by Gabara (1988) that a backside contact with very low impedance may help in mitigating the problem. This concept was applied, for instance, in the SVX3 front-end circuit described by Zimmermann et al. (1998). Many mixed-mode front-end ASICs successfully in use are fabricated on epi wafers using 0.35–0.25-μm CMOS processes. In a lightly doped substrate, on the other hand, currents tend to propagate more superficially. In this case, physical separation and guard rings are more helpful. Aragonès and Rubio (1999) found that the same test circuit fabricated on both types of wafers has a significantly lower noise when implemented on a lightly doped substrate.

Modern very-deep-submicron processes offer the possibility of having triple-well nMOS transistors. In these devices, a deep n-well is fabricated in the substrate. Inside the n-well, a p-well is formed to host the nMOS device. This has the advantage that the transistor can be very effectively insulated from the substrate. Furthermore, the source and the bulk can always be connected, thereby suppressing the body effect. Although this configuration takes significantly more layout area, triple-well nMOS is a natural candidate to implement very critical devices like the input stage of front-end amplifiers.

Several techniques for minimizing substrate noise have been proposed, some of which are rather elaborate (Peng and Lee 2004). Of course, the degree of sophistication depends on the requirements of the applications. We have a few basic suggestions.

1. Each analog block should have its own line for power, ground, and substrate biasing. In particular, it is important that the input stage of the analog front end does not share any of these lines with other circuits. Up to now, we have been concerned about noise injected by digital elements. However, for low-power applications, it can be helpful to use class AB analog circuits to reduce the power consumption. A common example is the output stage of a fast shaper that has to deliver relatively large signals to the rest of the processing chain. Since a class AB circuit can drive a signal current much big-

ger than the one required at the quiescent bias point, the current variation in its power-supply lines can be significant, and sharing those lines with the first amplifying stage can be detrimental.
2. Analog and digital blocks should be separated as much as possible and shielded by guard rings. Dedicated insulation trenches must be used if available in the process.
3. A careful pad assignment should be made, and enough pads should be reserved for the power and ground lines to minimize the series inductance in the connection.
4. If possible, one should use a digital library in which the bulk of the transistors are not tied to their source. Independent lines should be used to bias the substrate of the cells, and of course these lines should not be shared with their counterparts employed in the analog section.
5. Enough decoupling capacitors should be foreseen. Decoupling capacitors should also possibly be integrated on the chip.
6. The minimum value of the power-supply voltage that allows a correct functioning of the digital logic should be used.
7. Most of the switching activity occurs at the I/O ports, which should not have the power lines in common with the core logic located inside the chip. Low-voltage differential signaling (LVDS), which works by steering a constant current between two branches, should always be used, and CMOS pads should be avoided.

The use of LVDS often raises some concerns. On the one hand, an LVDS link requires twice the number of lines with respect to its CMOS counterpart. On the other hand, an LVDS driver steers a constant current of 2–4 mA; thus its power dissipation easily exceeds 10 mW per port, and it is independent of the frequency. To dissipate 10 mW, a CMOS driver working at 2.5 V should send a signal at 200-MHz frequency on an 8-pF load. However, the benefits provided by LVDS extend beyond the minimization of the switching noise on the chip. Its use is always recommendable and becomes a must whenever off-chip data transmission is concurrent with signal acquisition from the sensor. To better understand this point, consider the following example.

A front-end ASIC is specified to provide an ENC of 500 electrons rms. When the chip has a hit, it raises a flag to the data-acquisition system through a CMOS output port with a 2.5-V swing. A 0.5-fF parasitic capacitance on the PCB between the flag line and the input of the front-end amplifier determines the injection of a charge of 1.25 fC (7800 electrons), which is well beyond the specified noise limit. Even worse, if the signal exceeds the specified threshold, it may trigger another switching event, leading to a disastrous positive feedback. In the design of a mixed-mode system, one is often concerned by the interference coming from the clock. However, if the bandwidth of the front-end stage is small enough, the leading and trailing edges of the clock will occur before the front-end amplifier has the time to develop a full signal, and they will tend to cancel each other. Furthermore, if the amplifier output is sampled synchronously with the clock, the perturbation on the baseline will be captured with the same phase and will tend to manifest itself as an offset. However,

signals occurring asynchronously and lasting several clock periods can be much more harmful, since they give the front end the time to develop a full response.

Good mixed-mode practice generally tends toward an increase in the number of I/O pads and external decoupling capacitors. This leads, of course, to some complications in the development of the front-end board and to additional costs, which can be critical for some applications. The chip designer should nevertheless be prepared to resist the pressure to simplify the design that sometimes comes from other parties involved in a complex project, since a badly performing front end compromises the quality of the measurements and is often at the source of endless debugging efforts.

### 9.4.2 Common Mode Noise

The baseline levels of different channels are often observed to fluctuate coherently in time, superimposing a parasitic component, the common mode noise, on a genuine signal. One case that may happen is when the switching noise modulates the reference ground of the front-end amplifier. As illustrated in Figure 9.2, this disturbance can be amplified by the ratio between the parasitic capacitance connected to the input and the feedback capacitance.

Another situation occurs when data are transmitted outside the front-end chip via a single-ended link (Spieler 2006). In Figure 9.10a, a single-ended transmitter sends a signal to a receiver. The current flows back to the receiver via the common return path. Due to the finite impedance, a voltage drop may develop and contaminate the transmitted signal. Figure 9.10b depicts an LVDS-like connection, in which a current is steered in a well-controlled loop. The receiver senses only the difference between the two lines, thereby rejecting any disturbance that may affect both of them simultaneously. Furthermore, the return to the receiver occurs via a well-defined path,

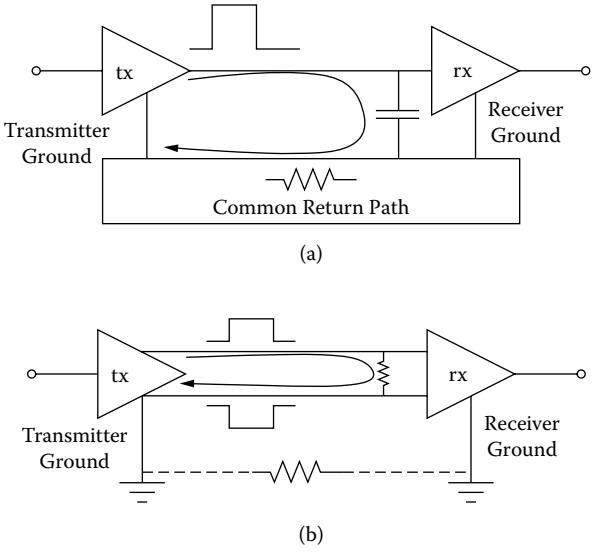



**FIGURE 9.10** (a) Single-ended and (b) differential signal transmission.

and the "grounds" of the two circuits need just a loose coupling, represented here by a resistor. If the receiving circuit is a noisy processor, it will be more difficult for its switching noise to reach the sensitive front end. In addition, having the currents circulating in well-controlled loops also minimizes the risk of unwanted pickups. Needless to say, these considerations apply independently of the form (analog or digital) in which data are transmitted.

Since common-mode noise affects many channels simultaneously, it can be calculated and subtracted. This can be done either off-line during data analysis or online with dedicated signal-processing stages. In self-triggering systems, if the baseline and the threshold are not equally affected, false hits can be flagged, and one would be forced to work with a higher threshold and a reduced sensitivity. In some cases, the threshold voltage is provided via a replica of the front-end amplifier, with the idea that common mode will appear on both inputs of the comparator and be rejected by its differential input stage. However, a radiation sensor is intrinsically single-ended, and it can be difficult to achieve an adequate match between the conditions at the input of a real channel and of the replica circuit. Thus, in the case that a mechanism such as the one depicted in Figure 9.2 is at play, this solution may not be completely effective.

## REFERENCES

Adolphsen, C., R. G. Jacobsen, V. Lüth, G. Gratta, L. Labarga, A. M. Litke, A. S. Schwarz, M. Turala, C. Zaccardelli, A. Breakstone, C. J. Kenney, S. I. Parker, B. A. Barnett, P. Dauncey, D. C. Drewer, and J. A. J. Matthews. 1992. The Mark II silicon strip vertex detector. *Nucl. Instr. Meth. A* 313:63.

Anelli, G., et al. 2001. A large dynamic range radiation-tolerant analog memory in a quarter-micron CMOS technology. *IEEE Trans. Nucl. Sci.* 48:435–439.

Anghinolfi, F., et al. 1995. A low-power piecewise linear analog to digital converter for use in particle tracking. *IEEE Trans. Nucl. Sci.* 42:772–775.

Aragonès, X., and A. Rubio. 1999. Experimental comparison for substrate noise coupling using different wafer types. *IEEE J. Solid-State Circuits* 34:1405–1409.

Badaroglu, M., et al. 2004. Impact of technology scaling on substrate noise generation mechanisms. In *Proc. IEEE Custom Integrated Circuits Conf. (CICC '04)*. Washington, D.C.: IEEE, 501–504.

Baur, R., et al. 1998. Front-end electronics for the CERES TPC-detector. *Nucl. Instr. Meth. A* 409:278–285.

Blalack, T., and B. A. Wooley. 1995. The effects of switching noise on an oversampling A/D converter. In *IEEE ISSCC Dig. Tech. Papers*. Washington, D.C.: IEEE, 274–275.

Bonvicini, V., et al. 2007. CASIS 1.0: A prototype VLSI front-end ASIC with ultra-large dynamic range and integrated ADC for silicon calorimetry in space experiments. *Nucl. Instr. Meth. A* 572:340–344.

Bosch, E., et al. 2003. The ALTRO chip: A 16-channel A/D converter and digital processor for gas detectors. *IEEE Trans. Nucl. Sci.* 50:2460–2469.

De Geronimo, G. 2000. A CMOS baseline holder (BLH) for readout ASICs. *IEEE Trans. Nucl. Sci.* 47:1857–1867.

De Geronimo, G., et al. 2002a. Analog CMOS peak detect and hold circuits, Part 1: Analysis of the classical configuration. *Nucl. Instr. Meth. A* 484:533–543.

De Geronimo, G., et al. 2002b. Analog CMOS peak detect and hold circuits, Part 2: The two-phase offset-free and derandomizing configuration. *Nucl. Instr. Meth. A* 484:544–556.

De Geronimo, G., et al. 2002c. Analog peak detector and derandomizer for high-rate spectroscopy. *IEEE Trans. Nucl. Sci.* 49:1769–1773.

Gabara, T. 1988. Reduced ground bounce and improved latch up suppression through substrate conduction. *IEEE J. Solid-State Circuits* 23:1224–1232.

Gatti, E., and P. Rehak. 1984. Semiconductor drift chambers: An application of a novel charge transport scheme. *Nucl. Instr. Meth. A* 253:365–377.

Haller, G. M., and B. A. Wooley. 1994. A 700-MHz switched-capacitor analog waveform sampling circuit. *IEEE J. Solid-State Circuits* 29:500–508.

Lan, H., et al. 2006. Synthesized compact models and experimental verifications for substrate noise coupling in mixed-signal ICs. *IEEE J. Solid-State Circuits* 41:1817–1829.

Mazza, G., A. Rivetti, G. Anelli, F. Anghinolfi, M. I. Martinez, and F. Rotondo. 2004. A 32-channel, 0.25 mm CMOS ASIC for the readout of the silicon drift detectors of the ALICE experiment. *IEEE Trans. Nucl. Sci.* 51:1942–1947.

Minderico, G., et al. 2003. A CMOS low power, quad channel 12 bit, 40 MS/s pipelined ADC for applications in particle physics calorimetry. Paper presented at the 9th Workshop on Electronics for LHC Experiments, Amsterdam, The Netherlands, 88–91.

Mota, B., et al. 2000. Digital implementation of a tail cancellation filter for the time projection chamber of the ALICE experiment. Paper presented at the 6th Workshop on Electronics for LHC Experiments, Krakow, Poland, 164–168.

Nouais, D., et al. 2003. The ALICE silicon drift detector system. *Nucl. Instr. Meth. A* 501:119–125.

Peng, M. S., and H. Lee. 2004. Study of substrate noise and techniques for minimization. *IEEE J. Solid-State Circuits* 39:2080–2086.

Rashevski, A., et al. 2002. Large area silicon drift detector for the ALICE experiment. *Nucl. Instr. Meth. A* 485:54–60.

Raymond, M., et al. 2005. The MGPA electromagnetic calorimeter readout chip for CMS. *IEEE Trans. Nucl. Sci.* 52:756–763.

Razavi, B. 1994. *Principles of Data Conversion System Design*, Wiley-IEEE Press.

Rivetti, A., et al. 2001. A low-power 10-bit ADC in a 0.25-μm CMOS: Design considerations and test results. *IEEE Trans. Nucl. Sci.* 48:1225–1228.

Rivetti, A., et al. 2005. The front-end systems of the silicon drift detectors of ALICE. *Nucl. Instr. Meth. A* 541:267–273.

Rodriques, P., et al. 2009. Experimental characterization of the 192 Clear-PEM ASIC coupled to a multi-pixel APD readout of LYSO:Ce crystals. *Nucl. Instr. Meth. A* 598:802–814.

Snoeys, W., et al. 2000. Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip. *Nucl. Instr. Meth. A* 439:349–360.

Spieler, H. 2006. *Semiconductor detector systems*. Oxford, U.K.: Oxford University Press.

Su, D. K., M. J. Loinaz, S. Masui, and B. Wooley. 1993. Experimental results and modeling techniques for substrate noise and mixed-signal integrated circuits. *IEEE J. Solid-State Circuits* 28:420–430.

Tedja, S., et al. 1995. A CMOS low-noise and low-power charge sampling integrated circuit for capacitive detector/sensor interfaces. *IEEE J. Solid-State Circuits* 30:110–119.

Trindade, A., et al. 2008. An overview of the clear-PEM breast imaging scanner. Paper presented at the Nuclear Science Symposium and Medical Imaging Conference, Dresden, paper MR3-3.

van Valburg, J., and R. van de Plassche. 1992. An 8-b 650-MHz folding ADC. *IEEE J. Solid-State Circuits* 27:1662–1666.

Yeom, J. Y., et al. 2007. A waveform sampling front-end ASIC for readout of GSO/APD with DOI information. *Nucl. Instr. Meth. Physics Research A* 571:381–384.

# 10 Current- and Charge-Sensitive Signal Conditioning for Position Determination

*Sven Peter Bönisch*

**CONTENTS**

## 10.1 INTRODUCTION

Signal conditioning is essential in obtaining and improving spatial as well as energy information in radiation-detection systems. Due to the generally extremely low signal levels supplied by a radiation detector, there is a strong demand for low-noise, high-gain, unconditionally stable, and best-matched signal-conditioning components. State-of-the-art signal conditioning is still mainly based on analog devices. However, there is an ongoing trend to integrate discrete analog structures and to substitute more and more analog devices by digital methods and procedures. Although the basic principles of signal-conditioning procedures, either analog or digital, are well understood [13, 14], the ongoing progress in technology and performance also demands an improvement in the theoretical background.

This chapter is devoted to discussion of analog signal conditioning for position determination. It explains in detail basic definitions, design specifications, procedures, and principles of signal conditioning to be used in any type of radiation-detection systems. It is a major concern to explain the difference between current- and charge-sensitive signal conditioning. Their advantages and drawbacks will be discussed. The general method of signal shaping will be introduced and explained based on the example of Gaussian delay-time approximation.

A special focus is on the theoretical description of the noise of a signal amplifier connected to a detector-specific input impedance, aiming at the improvement of the signal-to-noise ratio for best spatial and energy resolution. In position-sensitive detector (PSD) systems, two charge/current-sensitive preamp inputs may be connected via a coupling impedance. Due to this coupling, the output noise increases significantly. The behavior of that kind of coupling-induced noise is investigated using a technique called *noise gain calculation*. The output noise correlation is analyzed analytically. Attained results are cross-checked via simulations and measurements. Conclusions are drawn to better understand system resolution estimation for coupled preamplifier designs. The outcome may guide the reader in selection, design, and optimization of the signal-conditioning chain for position-sensitive radiation detection systems.

## 10.2 SIGNAL CONDITIONING

The signal-conditioning chain in a detector system is used to convert the charge supplied by the detector into a voltage or current pulse to be acquired and processed by the data-acquisition system. The classical approach is a charge-sensitive signal-conditioning system (Figure 10.1). It consists of a charge-sensitive preamplifier (CSA) and a pulse-shaping device. In the classical approach, the pulse shaping

has to be designed using a series connection of a differentiator (high-pass) and an n-pole integrator (low-pass). The charge-sensitive preamplifier is used to integrate the input pulse, thus forming a charge-proportional voltage pulse. The output pulse has a rather long decay time (≈100 μs). To increase the counting rate of the system, the pulse decay time must be reduced by two orders of magnitude. This is achieved by differentiation of the signal. The second integrator stage is used to form a defined pulse shape (i.e., semi-Gaussian), mainly by increasing the rise time of the rising slope. This design has been used over decades in many radiation-detection systems. Its main advantage is that it has been extensively investigated and is thus theoretically well understood [17]. The charge-sensitive design is still the best choice for detector systems requiring excellent energy resolution.

A somewhat newer approach is the current-sensitive signal conditioning (Figure 10.2). It has a much simpler construction than the classical approach. It consists of a current-sensitive preamplifier or current-to-voltage converter (ISA) directly followed by the pulse shaper low-pass. The integration of the charge is accomplished mainly by the pulse shaping itself. This also means that the pulse-shaping time is somewhat sensitive to the charge formation and collection time of the detector. Due to the nonideal behavior of the current-sensitive preamplifier, the input pulse is partly integrated in the first stage, too. The current-sensitive signal-conditioning approach is the best choice for detector systems requiring highly linear position determination.

There are applications requiring both good energy resolution and excellent position linearity. For these applications, the signal-conditioning system supplies a fast current signal and a slower energy signal [18].

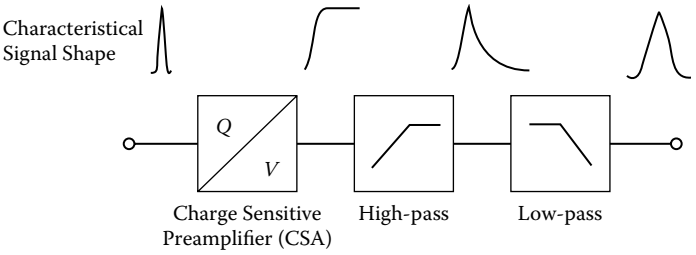



**FIGURE 10.1** Classical charge-sensitive signal conditioning.

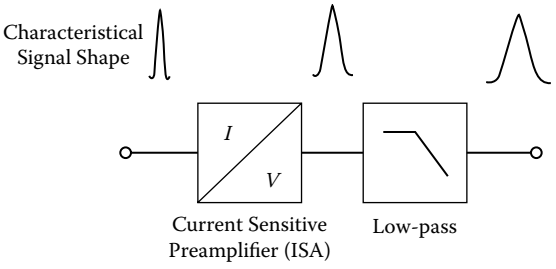



**FIGURE 10.2** Current-sensitive signal conditioning.

### 10.2.1 Preamplifier

The purpose of the preamplifier is to match the impedance of the detector into the signal-conditioning chain and to amplify the detector signal. It must provide signal amplification with extremely low additional noise, high linearity, and wide bandwidth. But typically all amplifiers are bandwidth limited, especially those incorporating feedback structures. A bandwidth limitation influences significantly the signal. Thus, the type of preamplifier can be classified by the relation of its own bandwidth to the signal bandwidth. If the preamp bandwidth is smaller than or equal to the signal bandwidth, it acts as a charge-sensitive amplifier. If the preamp supplies about one order of magnitude wider bandwidth, the output signal shape follows the input signal shape. Thus it acts as a current-to-voltage amplifier or as a voltage amplifier, depending on the feedback and detector impedance. Generally, all preamplifiers have low-pass characteristics.

#### 10.2.1.1 Charge-Sensitive Amplifier

The charge-sensitive amplifier (CSA) produces a charge-proportional output voltage. Therefore, it has to integrate the input current over a defined time. The charge-sensitive amplifier acts as an integrator in the time domain or as a low-pass in the frequency domain. The time dependency of the output voltage on the input current is given by

$$v_{\text{out}} \sim q_{\text{in}} = \int i_{\text{in}}(t) \cdot dt \tag{10.1}$$

A charge-sensitive amplifier (CSA) consists of an amplifier with negative feedback. The amplifier might be a transistor, an integrated circuit, or an operational amplifier. Here it is assumed that the amplifier is an operational amplifier. The general circuit is given in Figure 10.3. The operational amplifier has a quite high open-loop gain of about $g_{\text{OL}} = 10^3$–$10^5$ (60–100 dB). The feedback loop contains a

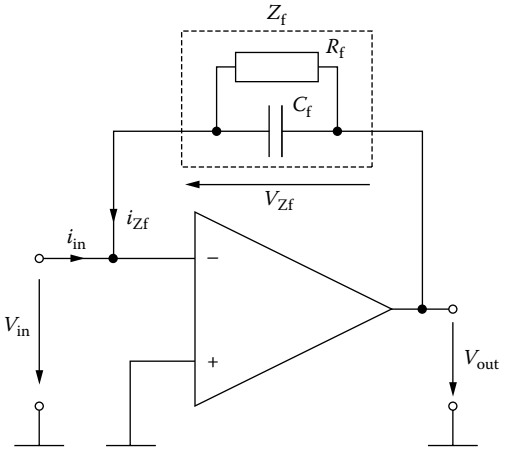



**FIGURE 10.3** Electric circuit of an ideal preamplifier stage (either CSA or ISA).

capacitor $C_f$ and a resistor $R_f$ in parallel connection forming the feedback impedance $Z_f$. The feedback resistance $R_f$ is on the order of a few megaohms (MΩ). The input signal is fed into the summing point of the inverting input and feedback. The gain of the CSA is given by

$$g_{Signal} = \frac{v_{out}}{i_{in}} = -Z_f = -X_{Cf} \left\| R_f = -\frac{1}{\frac{1}{R_f} + j\omega C_f} \right. \tag{10.2}$$

If the feedback resistance $R_f$ is assumed to be infinitely large, the charge gain in the time domain becomes $1/C_f$. Ideally, this device integrates immediately. In reality, the rise time of the output voltage is determined by the gain-bandwidth product of the amplifier. For a defined input charge, the output voltage is not stable. Due to the parallel connection of $R_f$ and $C_f$, it shows a decay time constant of

$$\tau_{decay} = R_f \cdot C_f \tag{10.3}$$

Thus only for a short time frame do we obtain the correct charge-proportional output voltage. This value must be sampled immediately after the end of the particle event. A time delay causes errors in energy and position determination due to the decay time constant and, in a coupled case, furthermore due to charge equalization [1].

The input impedance of a CSA is given by

$$Z_{\text{in}} = \frac{Z_f}{1 + g_{\text{OL}}} = \frac{1}{\frac{1}{R_f} + j\omega C_f} \cdot \frac{1}{1 + g_{\text{OL}}} \tag{10.4}$$

For high $R_f$ and $g_{\text{OL}}$ it reduces to

$$Z_{\text{in}} = \frac{1}{g_{\text{OL}} \cdot j\omega C_f} \tag{10.5}$$

yielding a so-called *Miller capacitance* seen at the input terminals. The equivalent Miller capacitance in the time domain is given by the feedback capacitance multiplied by the open-loop gain of the amplifier. Over the signal frequency range, the input impedance is purely reactive.

#### 10.2.1.2 Current-Sensitive Amplifier

The current-sensitive amplifier (ISA) produces a current-proportional output voltage. For detector impedances above the input impedance, it acts as a current-to-voltage converter. The shape of the output waveform follows the shape of the input waveform. This is mainly achieved by designing the ISA with a quite large signal bandwidth. The feedback time constant of a ISA should be at least 10 times below the smallest input current pulse width. The ISA shows the same device configuration as the CSA, except the feedback resistance is smaller by about three orders of magnitude (≈kΩ). The output voltage is proportional to the input current.

$$v_{\text{out}}(t) \propto i_{\text{in}}(t) \tag{10.6}$$

As the reactance of the feedback capacitance $X_{\text{Cf}}$ in the signal frequency range is much higher than the resistance of $R_{\text{f}}$, it can be neglected. The signal gain simplifies to

$$g_{\text{signal}} = \frac{v_{\text{out}}}{i_{\text{in}}} = -Z_{\text{f}} = -R_{\text{f}} \tag{10.7}$$

The gain of a voltage-to-current amplifier is sometimes referred to as *transimpedance*, given in units of volts per ampere (V/A).

The ISA shows no output signal decay time compared to the CSA.

The input impedance of the ISA, neglecting $X_{\text{Cf}}$, is given by

$$Z_{\text{in}} = \frac{R_f}{g_{OL}} \tag{10.8}$$

The input impedance of the ISA is determined by the ratio of feedback resistance and open-loop gain of the amplifier. The input impedance in the signal frequency range is purely resistive. It has no reactive or imaginary parts.

#### 10.2.1.3 Comparison of CSA and ISA

To compare both types of preamplifiers, the frequency- and time-domain behaviors need to be studied (Figures 10.4–10.6). For that purpose, simulations have been carried out using the electric circuit given in Figure 10.3. First of all, it can be clearly seen that both types of preamplifiers have low-pass characteristics. The only difference is that the ISA has a much wider bandwidth.

Typical input pulses having pulse lengths (Figure 10.6) in the submicrosecond range contain signal frequencies on the order of a few megahertz (MHz). Up to 1 MHz, the ISA has a flat frequency response (Figure 10.4). However, the gain of the CSA falls above the corner frequency given by the first feedback pole $1/(R_{\text{f}}C_{\text{f}})$ with −20 dB/decade. This is the reason for the integrative behavior. Both amplifier types CSA and ISA show in the frequency domain a passband gain (e.g., below the first-pole frequency) equal to the feedback resistance (transimpedance gain). The gain phase angle for both types in the passband is 180 degrees, they invert the input signal (Figure 10.4). Above the first-pole frequency an additional phase shift of −90° can be noticed. For the CSA, this is the reason for purely reactive input impedance in the signal frequency range.

The input impedance in the frequency domain (Figure 10.5) for signal frequencies shows similar behavior. Up to the first-pole frequency, the input impedance is purely resistive and determined by the ratio of feedback resistance to open-loop gain, falling by −20 dB/decade for frequencies above. Here an important note should be made: The input impedance of a CSA is well above zero and is always greater than the input impedance of the ISA! A common but serious misunderstanding is to see the CSA as an ideal current sink. The ISA is a much better current sink than the CSA.

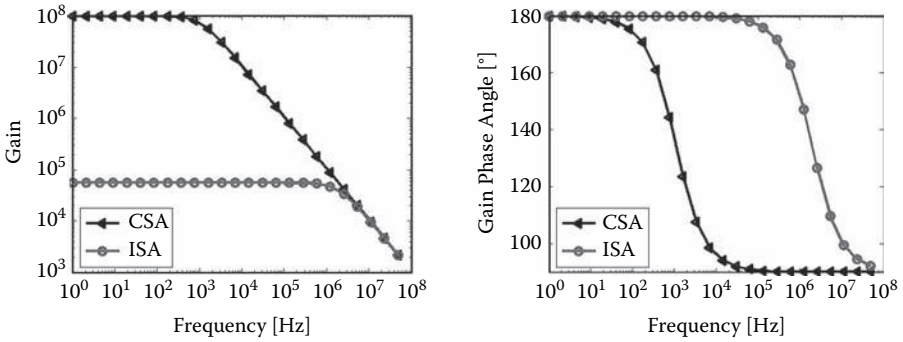



**FIGURE 10.4** Gain and phase of CSA and ISA in the frequency domain (CSA: $C_f = 1.5$ pF, $R_f = 100$ MΩ; ISA: $C_f = 1.5$ pF, $R_f = 56$ kΩ).

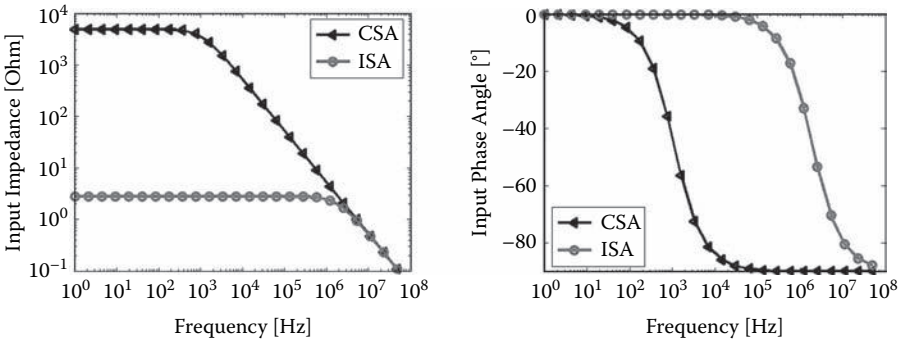



**FIGURE 10.5** Input impedance and phase of CSA and ISA (CSA: $C_f = 1.5$ pF, $R_f = 100$ MΩ; ISA: $C_f = 1.5$ pF, $R_f = 56$ kΩ).

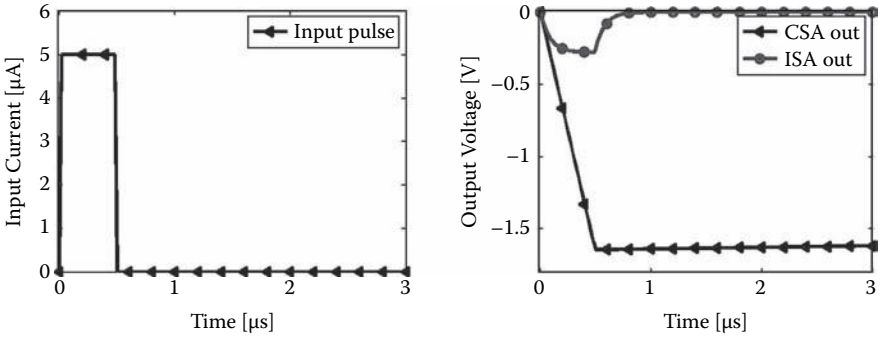



**FIGURE 10.6** Typical input signal and pulse response of CSA and ISA ($I_{max} = 5$ μA, $t = 500$ ns, $q_{in} = 2.5$ pC, single event; CSA: $C_f = 1.5$ pF, $R_f = 100$ MΩ; ISA: $C_f = 1.5$ pF, $R_f = 56$ kΩ).

Furthermore the CSA input impedance starts to become imaginary even for quite low frequencies (Figure 10.5). For the frequencies of interest, it is purely reactive.

In the time domain, both amplifier types invert the input signal. The CSA integrates the input signal with a conversion gain of $1/C_f$, while the ISA amplifies the signal by a transimpedance gain of $R_f$.

The output voltage for the CSA can be calculated by

$$v_{out}(t) = -\frac{1}{C_f} \cdot q_{in}(t) = -\frac{1}{C_f} \cdot \int_t i_{in}(t) \cdot dt \tag{10.9}$$

The output voltage for the ISA can be calculated by

$$v_{out}(t) = -R_f \cdot i_{in}(t) \tag{10.10}$$

The output voltage of the CSA falls later by the time constant, which is in the microsecond range. The nonideal behavior of the rising and falling slopes of the ISA is due to signal frequency components that fall partly into the integration region. This can be avoided by supplying preamp bandwidths that are more than one order of magnitude above the signal bandwidth.

#### 10.2.1.4 Real Preamplifier

In reality, detector systems contain additional components. In a more realistic scenario, we also have to consider (a) high-voltage blocking capacitors in the signal chain to block high voltage used for charge collection in the detector and (b) the detector impedance from the preamplifier input to ground (Figure 10.7). These

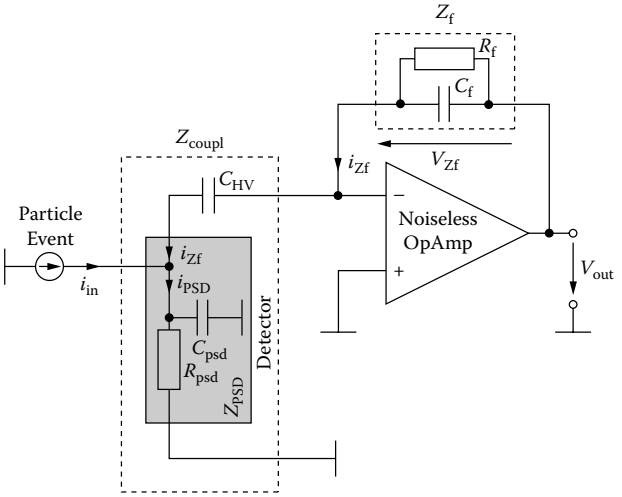



**FIGURE 10.7** Equivalent circuit of a realistic preamplifier stage used for signal gain and input impedance analysis (only one side shown, assuming ideal ground at the other side of the detector).

additional components influence the behavior of the preamplifier stage significantly. Furthermore, for position detection using the charge-division method, two preamplifier stages may be coupled via a coupling impedance (detector impedance) at their inputs (Figure 10.7). Given these additional factors, the signal gain and input impedance may differ from the ideal single circuits described previously. For derivation of the signal gain and input impedance of real preamplifiers, a circuit consisting of one ideal preamplifier, the optional high-voltage blocking capacitance $C_{\mathrm{HV}}$, and a PSD detector ($R_{\mathrm{PSD}}$, $C_{\mathrm{PSD}}$) grounded on the other side is assumed (Figure 10.7). A particle event supplies an input current $i_{\mathrm{in}}$ into a position very close to the upper end of the detector impedance. Under these assumptions, the signal gain is given by

$$g_{Signal} = \frac{v_{\mathrm{out}}}{i_{\mathrm{in}}} = -\frac{(X_{Cf} \| R_f) \cdot (X_{Cpsd} \| R_{psd})}{X_{CHV} + (X_{Cpsd} \| R_{psd})} \tag{10.11}$$

The input impedance of a realistic input stage is given by

$$Z_{\mathrm{in}} = Z_{\mathrm{PSD}} \| \left( X_{C_{\mathrm{HV}}} + Z_{\mathrm{in\,ideal}} \right) = \frac{Z_{\mathrm{PSD}} \cdot \left( X_{C_{\mathrm{HV}}} + \dfrac{Z_{\mathrm{f}}}{1 + g_{\mathrm{OL}}} \right)}{X_{C_{\mathrm{HV}}} + \dfrac{Z_{\mathrm{f}}}{1 + g_{\mathrm{OL}}} + Z_{\mathrm{PSD}}} \tag{10.12}$$

The analytical derivation of signal gain and input impedance reveals rather complicated terms. Their meaning will be discussed using an electric circuit simulation based on the circuit given in Figure 10.7.

The reactance $X_{\mathrm{CHV}}$ adds an additional zero in the complex S-plane. Thus realistic input stages exhibit a band-pass behavior, and in the lower frequency range their gain is limited by the reactance of the blocking capacitor $X_{\mathrm{CHV}}$ (Figure 10.8). They partly integrate and differentiate in the time domain. The signal gain in the passband region is further limited by the parallel detector impedance $Z_{\mathrm{PSD}}$. These realistic

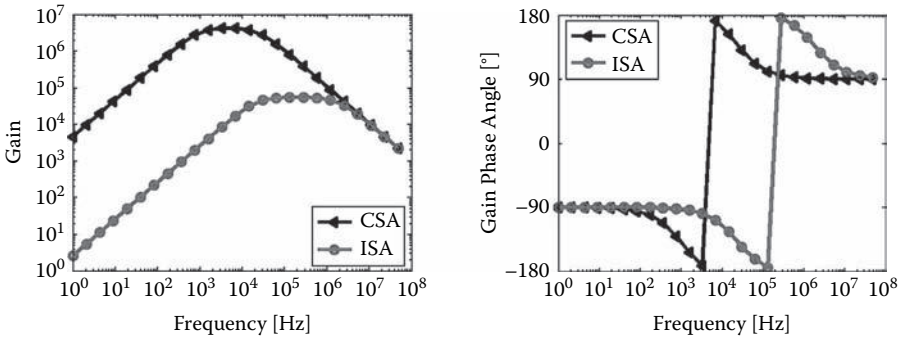



**FIGURE 10.8** Gain and phase of realistic preamplifier stages in frequency domain ($C_{\mathrm{HV}}$ = 3.3 nF, $R_{\mathrm{PSD}}$ = 2.2 kΩ, $C_{\mathrm{PSD}}$ = 15 pF; CSA: $C_{\mathrm{f}}$ = 1.5 pF, $R_{\mathrm{f}}$ = 100 MΩ; ISA: $C_{\mathrm{f}}$ = 1.5 pF, $R_{\mathrm{f}}$ = 56 kΩ).

preamplifier stages can transmit AC signals only; that is why they always introduce a baseline shift, which becomes significant for higher counting rates. The gain phase angle jumps from −90 degrees in the lower stop-band to +90 degrees in the upper stop-band (Figure 10.8). This means that the realistic input stage introduces a variable phase shift into the signal, depending on the input pulse length and current state of the preamplifier (time-dependent behavior will be discussed below). The additional zero frequency has to be considered in the pulse-shaper design.

For the given circuit, the input impedance is mainly determined by the impedance of the detector tied to ground at the other side (Figure 10.9). In the lower frequency range, the input impedance of the ideal CSA and ISA stages discussed previously is detached by the high reactance of the blocking capacitance $X_{CHV}$ connected in series to the input. In the higher frequency range, the input impedance is mainly determined by the $X_{CHV}$ roll-off (–20 dB/decade). Both types of preamplifiers show the same input impedance and input phase angle (Figure 10.9), and both have a complex input impedance (i.e., real and imaginary components).

More realistic scenarios must consider the time-domain response of the preamplifier stage on high and statistical distributed input event rates. Here we assume an evenly distributed 500-kHz event rate. The CSA configuration shows a steadily increasing output level, called *pulse pileup* (Figure 10.10). If the delay time between two successive input pulses is smaller than the first-pole time constant of the CSA, the output voltage cannot fall completely down to zero, and the next pulse is stacked on top of the preceding. When the maximum output voltage of the amplifier is exceeded, the results are energy and position errors. Another very important behavior has to be noted. The original decay time constant is shortened by about one order of magnitude. This is caused by the connection of any impedance (blocking capacitor, detector impedance) to the input of the CSA.

Second, due to the finite insulation resistance ($\leq 10^{12}$ Ω), the blocking capacitor has small but, in some applications, nevertheless significant DC leakage current. This leakage current may cause an output voltage shift (i.e., baseline shift) in applications using a CSA with high feedback resistance, as the DC leakage current is amplified in the nonintegrating frequency range (Figure 10.4).

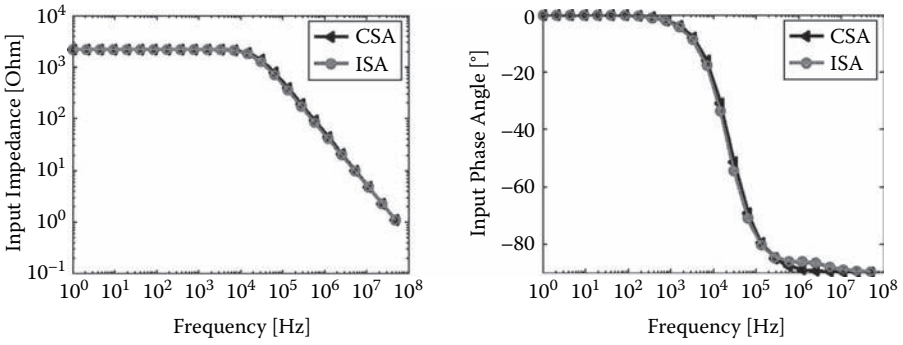



**FIGURE 10.9** Input impedance and phase of realistic preamplifier stages ($C_{HV}$ = 3.3 nF, $R_{PSD}$ = 2.2 kΩ, $C_{PSD}$ = 15 pF; CSA: $C_f$ = 1.5 pF, $R_f$ = 100 MΩ; ISA: $C_f$ = 1.5 pF, $R_f$ = 56 kΩ).

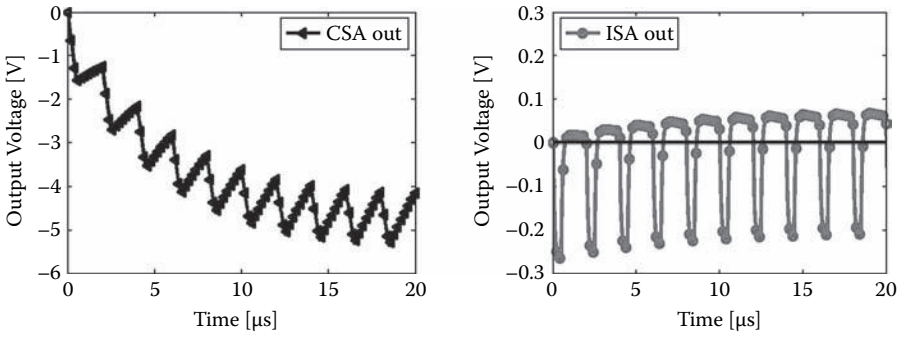



**FIGURE 10.10** Realistic preamp output signals for 500-kHz event rate ($C_{HV}$ = 3.3 nF, $R_{PSD}$ = 2.2 kΩ, $C_{PSD}$ = 15 pF; CSA: $C_f$ = 1.5 pF, $R_f$ = 100 MΩ; ISA: $C_f$ = 1.5 pF, $R_f$ = 56 kΩ; input signal: event rate = 500 kHz, $I_{max}$ = 5 µA, $t$ = 500 ns, $q_{in}$ = 2.5 pC).

The AC-coupled ISA configuration suffers a baseline shift too (Figure 10.10). This becomes evident at high counting rates, but the shift may be reduced using a baseline restorer later in the signal chain. Even at low counting rates, the blocking capacitance adds a zero in the complex S-plane. To obtain high accuracy, it has to be deleted by a pole of the same frequency in the pulse-shaping device.

## 10.2.2 Pulse Shaping

The second stage in the signal-conditioning chain is a pulse-shaping device. It is a special filter, designed to form a specific pulse shape in the time domain. A lot of information about common pulse-shaping techniques based on the classical CSA signal-conditioning approach can be found in the literature. This subject is not reviewed here. Instead, the reader is advised to seek out this information in the literature.

Often a semi-Gaussian pulse shape is required to apply analog-to-digital conversion (ADC) and further processing. One approach to form a semi-Gaussian pulse shape is the relatively unknown delay-time approximation. In conjunction with an ISA preamplifier stage, this is a low-pass having two tasks: (a) forming the desired pulse shape (i.e., semi-Gaussian) and (b) integrating the signal waveform supplied by the preamplifier stage to provide a nearly charge-proportional output signal.

### 10.2.2.1 Gaussian Delay-Time Approximation

There are various solutions to approximate an ideal Gaussian pulse at the output of a pulse-forming network. The easiest way to produce a semi-Gaussian pulse is to integrate the input signal $n$ times by simple integrators designed with the same time constant. The higher the network-grade $n$, the better the Gaussian pulse will be approximated by multiple real poles of the network function. In theory, the pulse response on $n$ integrators only lacks overshoots, but the use of real amplifiers with limited gain and frequency range often cannot avoid an additionally small ringing (i.e., over- and undershoots).

Better results can be expected by applying a method to specify a single delayed pulse in the time domain. This is the "Gaussian delay-time approximation," a

procedure described by Fritzsche [20]. The well-known standard multiple-real-pole approximation of a Gaussian transfer function in the frequency range minimizes errors in the frequency domain and obtains an optimum by varying only the real part. However, the approximation of the Gaussian delay-time filter includes the imaginary part and obtains the optimum by varying the real part of all complex poles by the same value. Because of that, it is possible to keep the errors of the approximation in the time domain occurring after the main pulse (i.e., over- and undershoots) as small as acceptable. The pulse response of the Gaussian delay-time filter shows only one overshoot (Figure 10.11). By specifying the height of overshoot $\delta_o$ in relation to the pulse maximum, the pulse-shaper structure can be fully determined.

By use of Equations (10.13) and (10.14), the poles of the transfer function of the Gaussian delay-time filter can be defined.

$$\omega_1 = 2\tan^{-1}\frac{2\nu\pi}{\ln\dfrac{1}{\delta_o}} \tag{10.13}$$

$$\sigma_1 = \frac{\omega_1}{2\pi}\ln\frac{1}{\delta_o} \tag{10.14}$$

The grade of the network function has been selected to $n = 5$, resulting in one real pole and two conjugate complex pole pairs (Equation [10.15]) and a transfer function (Equation [10.16]) prepared for cascade filter design.

Distribution of the poles:

$$\begin{aligned} P_1 &= -\sigma_1 \\ P_2 &= -\sigma_1 + j\cdot\omega_1 \\ P_3 &= -\sigma_1 + j\cdot 2\omega_1 \end{aligned} \tag{10.15}$$

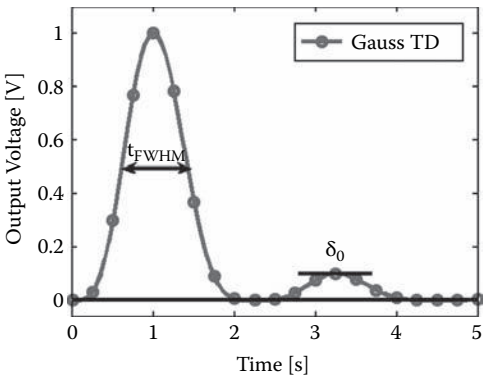



**FIGURE 10.11** Gaussian delay-time approximation in the time domain.

Transfer function:

$$H(S) = \frac{P_1}{(S-P_1)} \cdot \frac{P_2 P_2^*}{(S-P_2)(S-P_2^*)} \cdot \frac{P_3 P_3^*}{(S-P_3)(S-P_3^*)} \tag{10.16}$$

For practical use, an overshoot of 0.01% of the maximal pulse height should be sufficient. It corresponds to $\delta_o = 0.0001$ (Figure 10.11) and the following values for the locations of the poles.

Poles after initial design step:

$$\begin{aligned} P_1 &= -1.755 \\ P_2 &= -1.755 + \mathrm{j} \cdot 1.197 \\ P_3 &= -1.755 + \mathrm{j} \cdot 2.395 \end{aligned} \tag{10.17}$$

The pulse response after the initial design step is characterized by a nearly symmetrical shape without any visible overshoot, but with a shaping time $t_{\mathrm{FWHM}} = 1.56$ s. (FWHM is the abbreviation for full width at half maximum.) It is convenient for creating a common solution to normalize the shaping time $t_{\mathrm{FWHM}}$ to 1 s.

Poles after shaping-time normalization:

$$\begin{aligned} P_1 &= -2.742 \\ P_2 &= -2.742 + \mathrm{j} \cdot 1.871 \\ P_3 &= -2.742 + \mathrm{j} \cdot 3.742 \end{aligned} \tag{10.18}$$

The result of the Gaussian delay-time approximation is a more symmetrical and shorter pulse compared to the multiple-real-pole approximation assuming identical network degrees $n$ (Figure 10.12).

Gain and group delay plots complete the performance in the frequency domain (Figure 10.13).

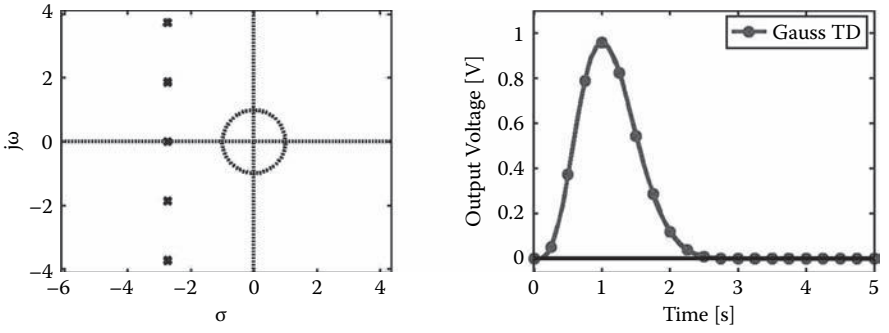



**FIGURE 10.12** Pole diagram and pulse response of the Gaussian delay-time filter (normalized).

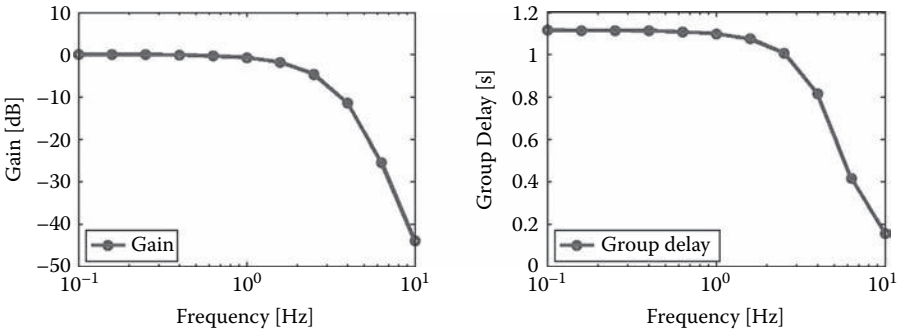



**FIGURE 10.13** Gain and group delay of the Gaussian delay-time filter (normalized).

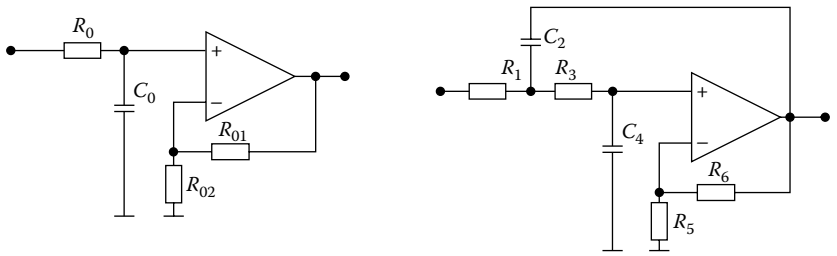



**FIGURE 10.14** First-order and second-order building blocks.

### 10.2.2.2 Building Blocks

To obtain the network transfer function in Equation (10.19), it is convenient to divide it into three sections, a first-order building block realizing the single pole and two consecutive second-order building blocks (Figure 10.14).

$$H(S) = \frac{2.742}{(S+2.742)} \cdot \frac{11.056}{(S^2+5.484 \cdot S+11.056)} \cdot \frac{21.59}{(S^2+5.484 \cdot S+21.59)} \tag{10.19}$$

Starting with the first-order building block and the corresponding real-pole equation (Equation 10.20), $R_0$ and $C_0$ can be calculated by comparison of the coefficients. If the time $t_{FWHM}$ is specified to 1 µs, all terms of frequencies must be divided by 1 µs. The following results can be obtained:

Assumptions: $\delta_o = 0.0001$, $t_{FWHM} = 1$ µs
Result for real pole: $C_0 = 470$ pF at $R_0 = 777\ \Omega$

$R_{01}$ and $R_{02}$ form an additional gain factor GF.

$$H(S) = \mathrm{GF} \cdot \frac{2.742}{(S+2.742)} = \frac{R_{01}+R_{02}}{R_{02}} \cdot \frac{1/C_0}{(S+1/C_0R_0)} \tag{10.20}$$

**TABLE 10.1**
**Element Values of the Second-Order Building Blocks**

| Elements | Second Building Block | Third Building Block |
|---|---|---|
| Resistor $R_1$ | 564 ohms | 442 ohms |
| Resistor $R_3$ | 730 ohms | 476 ohms |
| Resistor $R_5$ | 2000 ohms | 2000 ohms |
| Resistor $R_6$ | 827 ohms | 1698 ohms |
| Capacitor $C_2$ | 470 pF | 470 pF |
| Capacitor $C_4$ | 470 pF | 470 pF |

The calculation of a second-order building block is more complicated, as it is a predecessor that is introduced to minimize the gain-sensitivity product as a figure of merit to achieve low overall sensitivity. For detailed information, please refer to the literature [19]. As a result of this mainly software-based design procedure of second-order building blocks, all the missing-element values are tabulated in Table 10.1.

### 10.2.3 Example

The design process of an analog signal-conditioning system is characterized by trading off position error, energy error, and timing error. Sophisticated software tools, such as electric circuit simulators (APLAC, PSPICE, etc.) and mathematical software (MATLAB, MATHEMATICA, etc.), aid the design process. For a coupled system using the current signal-conditioning approach, almost no theoretical background exists. Thus, the design process is based on software tools only.

In this section, an analog signal-conditioning design example for a position-sensitive neutron detector is described at the circuit level. The design specifications for the detector system are:

- Counting rate ≥100 kHz per channel
- Timing resolution ≤100 ns
- Position resolution ≤1%
- Differential nonlinearity ≤2%
- Position deviation ≤1 mm
- Dark counting rate ≤0.2 counts/h

The electrical circuit diagram of the complete analog signal-conditioning chain is shown in Figure 10.15. The current pulse produced by a particle event is fed via a 15 nF/6 kV high-voltage coupling capacitor and a 270 Ω extension resistance into a current-to-voltage converter with integrated ESD (electrostatic discharge) protection diodes. The amplification factor is given by $R_f = 56$ k (i.e., $g = 56 \times 10^3$ V/A) for pulse lengths above the pole time constant of 56 ns. For the detector used in this system, this condition is fulfilled. Investigations have clearly shown that the preamplifiers must be configured as current-to-voltage converters to achieve a differential nonlinearity below 2%. After that, a pulse shaping is applied using a 1 μs FWHM (426-ns sigma), fifth-order Gaussian delay-time approximation with 0.01%

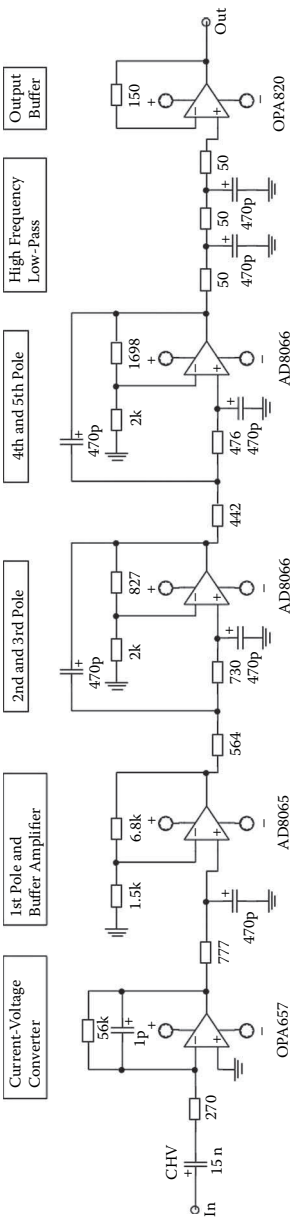



**FIGURE 10.15** Electrical circuit diagram of the analog signal conditioning (one channel).

undershoot (Equation [10.21]). Furthermore, the shaping circuit has been optimized for the lowest gain-sensitivity product (GSP = 1.065, from Moschytz and Horn [19]) to obtain reliable pulse response and best channel matching. The output buffer acts as a line driver for about a 2-m length of 50-Ω coaxial cable.

$$H(S) = k_1 \cdot \frac{2.738 \cdot 10^6}{(S - P_1)} \cdot \frac{237.2 \cdot 10^{24}}{(S - P_2)(S - P_2^*)(S - P_3)(S - P_3^*)} \tag{10.21}$$

$$k_1 = 5.533$$

$$P_1 = -2.742 \cdot 10^6$$

$$P_2 = -2.742 \cdot 10^6 + \mathrm{j} \cdot 1.871 \cdot 10^6$$

$$P_3 = -2.742 \cdot 10^6 + \mathrm{j} \cdot 3.742 \cdot 10^6$$

## 10.3 DESIGN SPECIFICATIONS

### 10.3.1 Conversion Gain

The conversion gain of a signal-conditioning system is defined as the relation of peak output voltage to peak input signal. In the case of a charge-sensitive system, it is the relation of peak output voltage to input charge, often given in units of volts per picocoulomb (V/pC). Due to the extremely low detector signals, the conversion gains for real signal-conditioning systems are on the order of 1 V/pC.

$$g_{\mathrm{conv}} = \frac{v_{\mathrm{out}}}{q_{\mathrm{in}}} \left[ \frac{V}{C} \right] \tag{10.22}$$

In the case of a current-sensitive system, it is the relation of peak output voltage to peak input current.

$$g_{\mathrm{conv}} = \frac{v_{\mathrm{out}}}{i_{\mathrm{in}}} \left[ \frac{V}{A} \right] \tag{10.23}$$

In a system using an ISA preamp, the shaper is partly or even fully integrating the input pulse; consequently, the conversion gain of this system type may also be given in units of volts per picocoulomb (V/pC).

### 10.3.2 Dynamic Range

The signal-conditioning chain, or generally any amplifier, can only be used within a defined signal-amplitude range (Figure 10.16). The lower limit is the noise level, and signals below the noise level are undetectable. The output of the amplifier shows its own amplified input noise. The upper hard limit is the maximum output voltage. Rail-to-rail amplifiers allow maximum output levels close to the supply voltage. However,

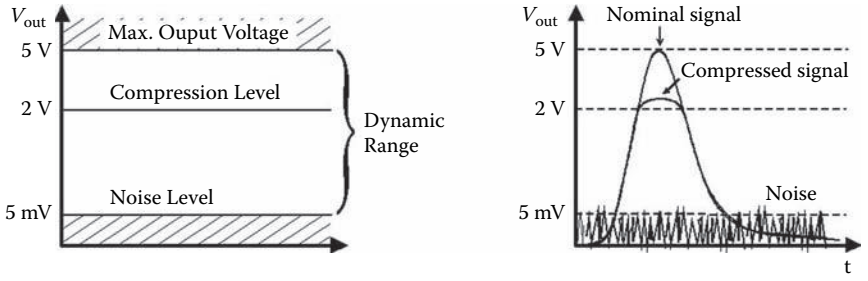



**FIGURE 10.16** Dynamic range and compression levels of an amplifier.

there is also a lower soft limit. As the output signal of the amplifier increases, the output stages start to compress the signal before reaching the maximum output voltage level. The effect is a decreased gain for higher signal amplitudes, causing signal distortion and, in the worst case, signal clipping and linearity errors (i.e., position and energy errors). For a highly linear operation, the amplifier should be used within a lower range than that suggested by its full dynamic range. The dynamic range is a unitless quantity, although it is often given in decibels (Equation 10.24). State-of-the-art operational amplifiers provide dynamic ranges on the order of 60–100 dB. For detector systems requiring a specific energy or position resolution, the gain compression has to be in the same order (i.e., 1% position resolution requires as low as 1% gain compression at the upper dynamic range).

$$D = 20 \cdot \log \frac{V_{\text{out\_max}}}{V_{\text{noise}}} [\text{dB}] \tag{10.24}$$

### 10.3.3 Linearity

#### 10.3.3.1 Signal Linearity

As the output signal approaches the maximum output voltage, it is distorted by compression and, in the worst case, signal clipping may occur. A common definition of signal distortion is the total harmonic distortion (THD). Unfortunately, the definition assumes harmonic signals in the frequency domain. Considering this, the total harmonic distortion is defined as the power sum of all harmonics normalized to the power of the fundamental frequency itself. Because the squares of the voltage are related to the signal power, we define THD as

$$\text{THD} = 20 \cdot \log \frac{\sqrt{\sum_{n} V^2_{n\text{th harmonic}}}}{V_{\text{fundamental}}} [\text{dB}] \tag{10.25}$$

Newer operational amplifiers and ADCs are specified by means of the spurious free dynamic range (SFDR) expressed in units of decibels to carrier (dBc). This definition considers only the highest (often the first) harmonic.

$$\mathrm{SFDR} = 20 \cdot \log \frac{V_{\text{fundamental}}}{V_{\text{highest harmonic}}} \left[\mathrm{dBc}\right] \tag{10.26}$$

The THD and the SFDR can be transferred into the time domain by means of a Fourier analysis. Sophisticated design tools (i.e., electric circuit simulators) and the use of circuit simulation at the transistor level perform this automatically. However, macromodels for newer operational amplifiers (OpAmps) often do not predict distortion. In this case, compression and distortion effects have to be investigated by measurements. For high-resolution detector systems, any distortion for frequencies of interest should be well below −60 dB.

### 10.3.3.2 Position Linearity

#### *10.3.3.2.1 Differential Nonlinearity*

Another, but physically completely different, type of linearity can be defined with respect to a position-sensitive detector system. An ideal position response of a detector system shows a linear dependency of real to measured position (Figure 10.17). However, due to different physical causes, the curve slope may change over the active detector area. The nonideal slope change between adjacent positions or statistical containers (i.e., channels of a spectrum) is defined as a differential nonlinearity (DNL). It is determined by the ratio of the complete differentials of real position (dPosition_real) and measured position (dPosition_measured) (Equation 10.36).

$$\mathrm{DNL} = \frac{\mathrm{dPosition\_real}}{\mathrm{dPosition\_measured}} \tag{10.27}$$

If only one coordinate (i.e., *x*-direction) is detected, this reduces to

$$\mathrm{DNL} = \frac{dx_{\text{real}}}{dx_{\text{measured}}} \left[\frac{\mathrm{mm}}{\mathrm{ch}}\right] \tag{10.28}$$

often given in units of millimeters per channel.

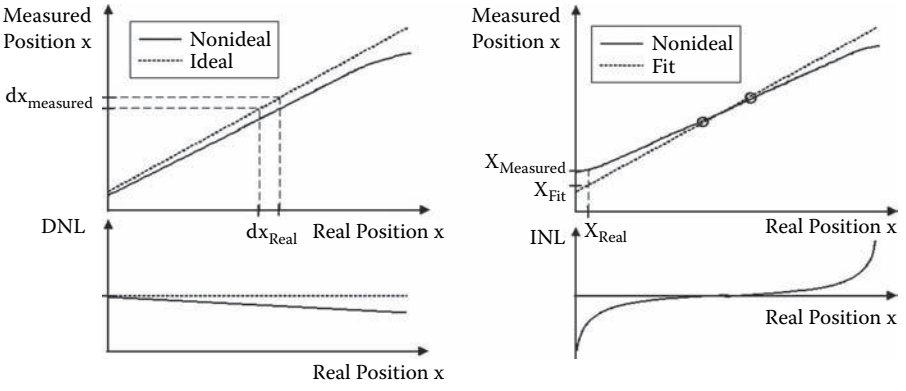



**FIGURE 10.17** Differential and integral nonlinearity of the position response.

#### 10.3.3.2.2 Integral Nonlinearity

More practical, however, is the integral nonlinearity (INL) over the full range of the detector. The nonideal position response is fitted using two points left and right of the curve center (i.e., middle of the response trace) by an ideal straight line (i.e., linear function) (Figure 10.17). The difference between the ideal fitting curve and the nonideal response trace determines the system's integral nonlinearity or position error (sometimes also referred as the deviation from nominal position), as seen in Equation 10.29. It may be given in units of either channels or mm.

$$\text{INL} = \text{Position\_fit} - \text{Position\_measured} \tag{10.29}$$

or

$$\text{INL} = x_{\text{fit}} - x_{\text{measured}}\,[\text{ch}], [\text{mm}] \tag{10.30}$$

Both linearity definitions describe the same physical reality. They are a measure for the nonideal behavior of the system position response.

#### 10.3.3.2.3 Linearity Correction

Investigations have shown that the differential nonlinearity of the position is the most critical design parameter in a coupled signal-conditioning system utilizing the charge-division method. The differential nonlinearity depends on many parameters (listed with descending significance):

1. Preamplifier (ISA)
2. Shaping time
3. HV-blocking capacitance
4. Input impedance (including additionally series resistors)
5. Detector time constant
6. Channel gain matching

The least significant, but nevertheless important, parameter is the gain matching of both analog signal-conditioning channels at either end of a detector. A gain mismatch of about 1% causes about a 5% linearity error or even more [16]. To overcome that problem, a digital or analog gain channel correction scheme can be used to decrease linearity errors (Figure 10.18).

As it is quite difficult to adjust the gain in the analog signal chain itself, the gain correction may be implemented on the software level, where a constant is multiplied by the measured peak value to achieve best channel gain matching. The gain correction factor can be calculated using the ideal position response symmetry [16]. The positive solution of a square function after application of the position calculation is the gain correction factor for channels 1 or 2, as seen in Equation (10.31). The relative error of this gain correction method for statistical signals is less than 0.024% (>12-bit accuracy).

$$c = \sqrt{\frac{V_{1\_\text{left}} \cdot V_{1\_\text{right}}}{V_{2\_\text{left}} \cdot V_{2\_\text{right}}}} \tag{10.31}$$

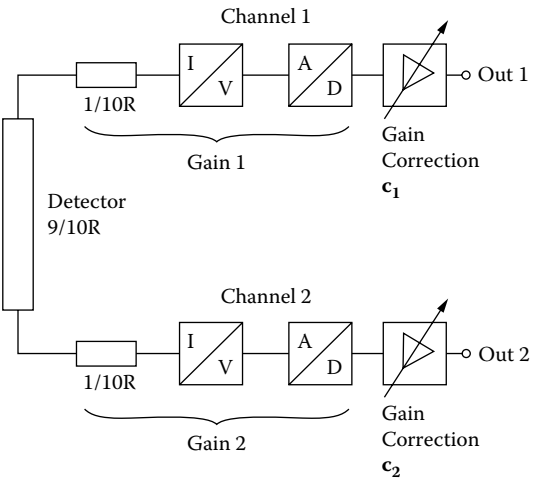



**FIGURE 10.18** Schematic of the digital gain correction.

### 10.3.4 Equivalent Noise Charge

The equivalent noise charge (ENC) is a common definition of the noise performance of a signal-conditioning system. It maps all noise sources in the signal chain to an equivalent input noise source. This could be either a noise voltage source or a noise current source. Both types of noise sources are convertible into each other. Here, a voltage noise source $V_N$ is assumed (Figure 10.19). The equivalent input noise voltage source is driving a noise current though the input capacitance and the attached detector capacitance $C_{PSD}$. This allows the definition of a charge, the ENC (Equation [10.32]) at the attached detector capacitance, as a measure of the system noise. It may be given in units of coulombs or the number of electrons.

$$\mathrm{ENC} = \left(C_{\mathrm{in}} + C_{\mathrm{PSD}}\right) \cdot V_{\mathrm{N}} \left[\mathrm{C}\right], \left[\mathrm{e}\right] \tag{10.32}$$

An increase in the detector capacitance or simply the connection of a detector causes a massive increase of the ENC, while the driving force, the equivalent input noise

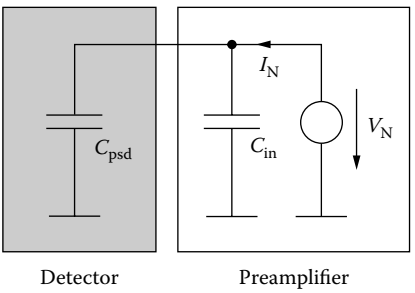



**FIGURE 10.19** Equivalent input circuit for the definition of the equivalent noise charge (ENC).

voltage, remains constant. The ENC in datasheets is often given without considering any detector capacitance.

A serious drawback is that this simple definition cannot correctly map the physical quantity "charge" from the input to the output of the signal-conditioning chain in the case of bandwidth limitations (i.e., real CSA or ISA preamplifiers, blocking capacitors, signal shaping) and especially in the case of differences between noise gain and signal gain (see Section 10.4). However, it is a well accepted measure of the noise performance of a signal-conditioning system.

### 10.3.5 Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) is closely related to the system dynamic range. It defines the relation of a signal to the system noise level. The definition may be referred to the input or output of the signal-conditioning chain; the physical quantities may be any of voltage, current, or charge; and it may be defined using peak or rms values. The ratio is often given in decibels.

$$\mathrm{SNR} = 20 \cdot \log \frac{\mathrm{Signal}}{\mathrm{Noise}} [\mathrm{dB}] \tag{10.33}$$

The SNR or dynamic range of a CSA is often defined by

$$\mathrm{SNR} = \frac{\text{maximum detectable charge}}{\mathrm{ENC}} \tag{10.34}$$

### 10.3.6 Position Resolution

The noise is closely related to the position and energy resolution of a detector system. In this section, only the behavior of the position resolution is discussed. Here, the rms (root mean square) value of the system output noise is equal to the standard deviation ($\sigma$) of a position measurement (Figure 10.20). This can be converted into a common definition of the position resolution at the full width at half maximum

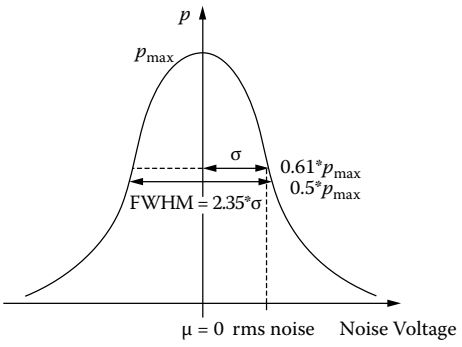



**FIGURE 10.20** Dependency of rms noise voltage and position resolution.

(FWHM) of the position distribution with a mean value μ and an output signal $Q_1$ (Equation 10.35). There are various definitions of the position resolution to be found in the literature. They should be applied with care.

$$\text{resolution FWHM} = \frac{2.35 \cdot V_{\text{Noise1 rms}}}{Q_1} \tag{10.35}$$

### 10.3.7 Shaping Time

The data-acquisition system demands a certain pulse length to obtain a correct reading (i.e., fewest energy and position errors). As explained in Section 10.2.2, the pulse shaping device forms the required pulse shape and a defined pulse width. Most often, a semi-Gaussian pulse shape is produced. Similar to the standard deviation σ of the Gaussian probability function, a shaping time σ can be defined at 61% of the peak value (Figure 10.21). This must not be confused with another common definition of the shaping time FWHM (i.e., pulse width) at 50% of the peak value. Both definitions can be transformed using Equation (10.36).

Sometimes the shaped pulse is characterized using the peaking time definition. The peaking time is defined as the time point at which a single pulse reaches its maximum (i.e., peak value). Assuming an ideal Gaussian pulse shape, it is about 3–5 times greater than the shaping time σ.

It should also be noted that all timing definitions given here are independent of the detector time constant only as long as the detector time constant itself (i.e., charge collection time, RC-time constant) is below or at least equal to the half of the pulse peaking time.

$$\text{Shaping time FWHM} = 2.35 \cdot \text{Shaping time } \sigma \tag{10.36}$$

$$\text{Peaking time} \approx 3 - 5 \cdot \text{Shaping time } \sigma \tag{10.37}$$

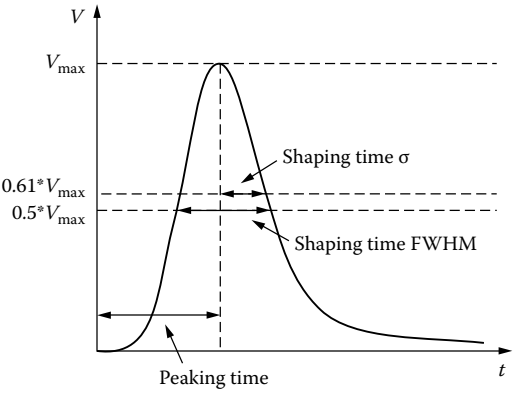



**FIGURE 10.21** Shaping time definitions.

## 10.4 NOISE

Noise in detector systems using the classical charge-sensitive signal-conditioning approach has been the subject of many published studies [1–8]. In most applications, noise has been referred to an equivalent input-noise source coming from that. A specification of the input referred noise, known as equivalent-noise charge (ENC), has been established. From ENC, a commonly accepted system resolution definition has been derived ($2.35 \times \text{ENC}/Q_{\text{signal}}$) [3]. This resolution estimation relies on the equality of signal gain and noise gain. In some applications of position-sensitive detector (PSD) systems utilizing the charge-division method, signal gain and noise gain may be different.

In classical charge-division PSD systems, two charge-sensitive preamplifiers are connected via a coupling impedance. This device structure exhibits an excessive low-frequency noise [1], due to an increase of the noise gain of the coupled circuit. The excessive low-frequency noise is much higher than all other noise sources considered so far [4] and does not follow the rules given. Thus, for correct estimation of the system resolution and to accommodate the difference between signal gain and noise gain, the noise gain must be taken into account.

This section describes the theoretical background of a coupled noise analysis and a noise correlation analysis for resistively coupled charge-sensitive preamplifiers. The behavior of noise gain, output noise, and noise correlation in different frequency ranges will be discussed in detail. All results are cross-checked via electrical simulations and measurements. Conclusions are drawn to better understand the estimation of system resolution for coupled preamplifier designs.

### 10.4.1 Operational Amplifier Noise Modeling

It is common in operational amplifier (OpAmp) noise modeling to refer all circuit noise sources back to the inputs using equivalent input noise sources. This can be done because the noise of the first stage in a high-gain circuit contributes mostly to the overall circuit noise. An OpAmp noise model consists of an equivalent input (series) voltage noise source and an equivalent input (parallel) current noise source connected to the inputs of a noiseless OpAmp (Figure 10.22). The equivalent input noise is given in terms of a voltage noise density (V/√Hz) and a current noise density (A/√Hz). State-of-the-art OpAmps like the OPA657 feature as low as 4.8 nV/√Hz input voltage noise density and 1.2 fA/√Hz input current noise density at a gain band-

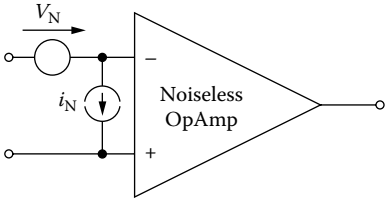



**FIGURE 10.22** Common OpAmp noise model with equivalent input noise sources.

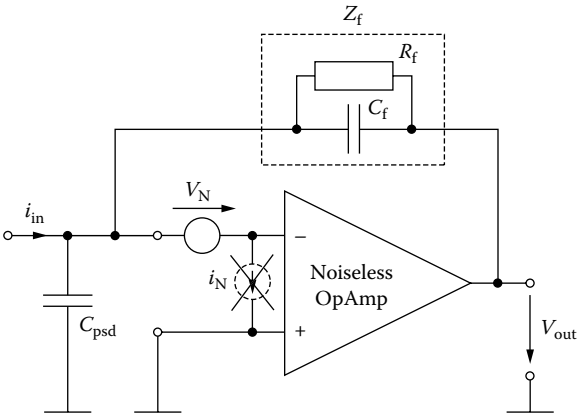



**FIGURE 10.23** Transimpedance amplifier design.

width product of 1.6 GHz. However, for analysis and simulation, the gain bandwidth product of the ideal, noiseless OpAmp is assumed to be infinite.

It is important to estimate the contribution of the voltage and current noise of the OpAmp equivalent input noise sources, including other noise sources (e.g., thermal noise of resistors) in the actual circuit [11, 12]. In the case of the classical wideband transimpedance design (Figure 10.23), which also covers the noise behavior of the charge amplifier, and using Equation (10.38), it can be shown that the equivalent input current noise of the OpAmp becomes negligible compared to the equivalent input voltage noise of the OpAmp [12].

$$i_{eq} = \sqrt{i_N^2 + \frac{4kT}{R_f} + \left(\frac{v_N}{R_f}\right)^2 + \frac{\left(2\pi C_{PSD} B_{lim} v_N\right)^2}{3}} \tag{10.38}$$

where

$i_{eq}$ = equivalent input noise current of the whole circuit with bandwidth limitation due to feedback
$v_N$ = OpAmp input noise voltage density (4.8 nV/√Hz)
$i_N$ = OpAmp input noise current density (1.2 fA/√Hz)
$R_f$ = feedback resistor (100 MΩ)
$C_f$ = feedback capacitor (1.5 pF)
$C_{HV}$ = HV blocking capacitor (3.3 nF)
$R_{PSD}$ = detector resistance (2.2 kΩ)
$C_{PSD}$ = detector capacitance to ground (15 pF)
$T$ = temperature (300 K)
$B_{lim}$ = bandwidth limitation frequency with feedback pole frequency
$B_{lim}$ = $1/(2\pi R_f C_f)$

Using the given values, the thermal noise current of the feedback resistor $R_f$ dominates the equivalent input-noise current $i_{eq}$, with about 13 fA/√Hz. The equivalent

input noise current $i_{eq}$ causes a voltage drop at $C_{PSD}$. In the classical transimpedance design (without $R_{PSD}$), equivalent input noise voltage and equivalent input noise current of the OpAmp become equal at about 28 kHz. Above 28 kHz, the contribution of the equivalent input noise current of the OpAmp to the overall circuit noise drops; for approximately $f > 280$ kHz, it becomes practically negligible.

### 10.4.2 Noise Gain

A classical position-sensitive detector system utilizing the charge division method is usually constructed of two sets of current/charge-sensitive preamps and shaping amplifiers, connected at the inputs via a detector-specific coupling impedance (Figure 10.24). These coupled systems suffer from excessive voltage noise in the low-frequency range [1]. The output noise increases significantly (more than 20 dB) below 5 MHz if two charge-sensitive preamp inputs are connected via a detector impedance. A lower impedance $R_{PSD}$ increases the output noise voltage. This excessive low-frequency noise is the main reason for limited system resolution [1]. The focus of this section is on the increased low-frequency noise below 5 MHz, which is mainly caused by the coupling of the two charge amplifiers via the detector impedance, whether or not the detector is assumed to be noisy. And, of course, the coupled second amplifier adds its noise too, but the increase of the noise due to the coupling is much more significant. In the case of a single-charge amplifier, the noise below 5 MHz is more than an order of magnitude less than in the coupled case.

To analyze a coupled system of two charge-sensitive amplifiers, we use the equivalent electrical circuit as shown in Figure 10.25. It consists of a noiseless position-sensitive detector impedance ($R_{PSD}$, $C_{PSD}$) and one noisy charge amplifier including an equivalent input noise voltage source. The second end of the detector is

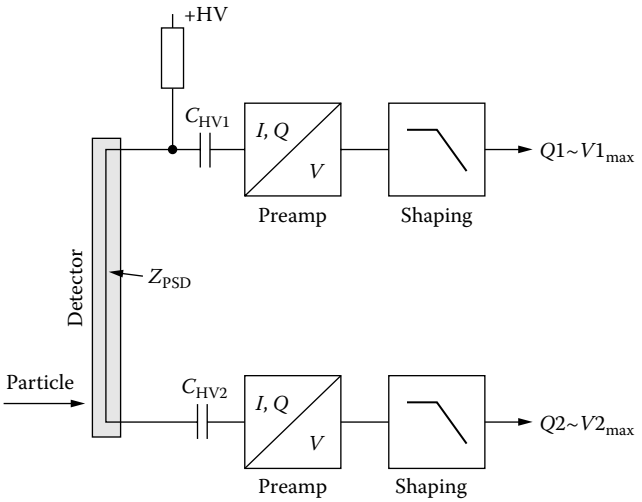



**FIGURE 10.24** Structure of PSD detector with preamplifiers (CSA or ISA) and shaping.

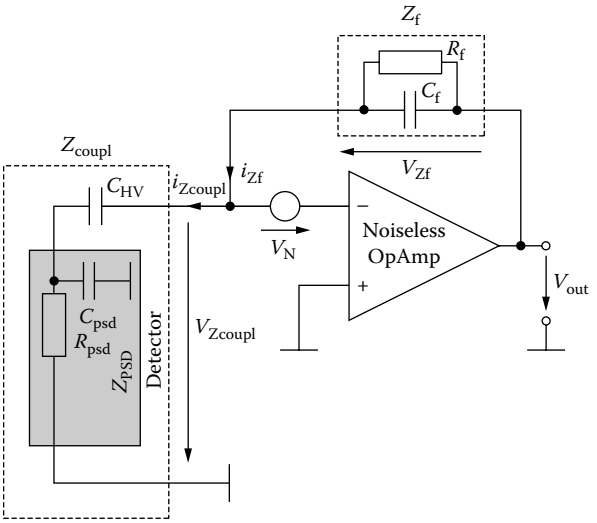



**FIGURE 10.25** Equivalent circuit for coupled noise gain analysis (one side of the detector shown, assuming noiseless $R_{PSD}$ and an ideal ground at the other side of the detector).

connected to ground, due to the assumption of an infinitely large coupling capacitance and an ideal noiseless charge amplifier input of 0 Ω.

The technique used for the noise analysis is called *noise gain calculation* [11]. It has never been considered in any publication dealing with coupled charge amplifiers in position-sensitive detectors. $R_f$ contributes to the circuit behavior mainly below the feedback pole frequency (1 kHz). That is why it is not considered for noise analysis.

The noise gain of the coupled circuit (Figure 10.25) is given by [15], Equation (10.39).

$$g_{noise} = 1 + \frac{X_{C_f} \parallel R_f}{X_{C_{HV}} + (X_{C_{PSD}} \parallel R_{PSD})} \tag{10.39}$$

For convenient understanding of noise gain and the output noise voltage behavior of such a coupled circuit, we need to consider at least three different frequency ranges (Table 10.2). Above 5 MHz (Range III), the noise gain is constant up to the OpAmp gain bandwidth limitation and will be determined by the ratio $C_{PSD}/C_f$. Thus, the output noise voltage density increases with the detector capacitance only within this range. At lower frequencies, the detector capacitance only influences the signal amplitude, not the noise. At around 5 MHz, $X_{CPSD}$ becomes equal to $R_{PSD}$. This point may also be referred to by the detector time constant $\tau_{PSD}$, where the cutoff frequency can be calculated by $1/(2\pi R_{PSD}C_{PSD})$. Below 5 MHz (Range II), the noise gain increases to lower frequencies, with $1/f$ depending on the ratio $X_{Cf}/R_{PSD}$. Below 20 kHz (Range I), we need to consider the blocking capacitance $C_{HV}$ as its reactance $X_{CHV}$ becomes comparable to $R_{PSD}$. In other words, the blocking capacitor $C_{HV}$ contributes mainly for times above the time constant $R_{PSD}C_{HV} \approx 7.3$ µs ($f < 20$ kHz). In this range, the noise gain is extremely high (>1000 for real devices) and determined by the ratio

**TABLE 10.2**
**Frequency Ranges and Noise Gain Approximation**

| Range | I | II | III |
|---|---|---|---|
| Frequency | <20 kHz | <5 MHz | >5 MHz |
| $X_{CHV}$ | $>R_{psd}$ | $<R_{psd}$ | $<R_{psd}$ |
| $X_{Cf}$ | $>R_{psd}$ | $>R_{psd}$ | $>R_{psd}$ |
| $X_{Cpsd}$ | $>R_{psd}$ | $>R_{psd}$ | $<R_{psd}$ |
| $g_{noise}$ | $\approx \frac{C_{HV}}{C_f}$ | $\approx \frac{X_{cf}}{R_{PSD}}$ | $\approx \frac{C_{PSD}}{C_f}$ |
| | 2200 | >12 | 11 |
| Slope | constant | $\approx 1/f$ | constant |

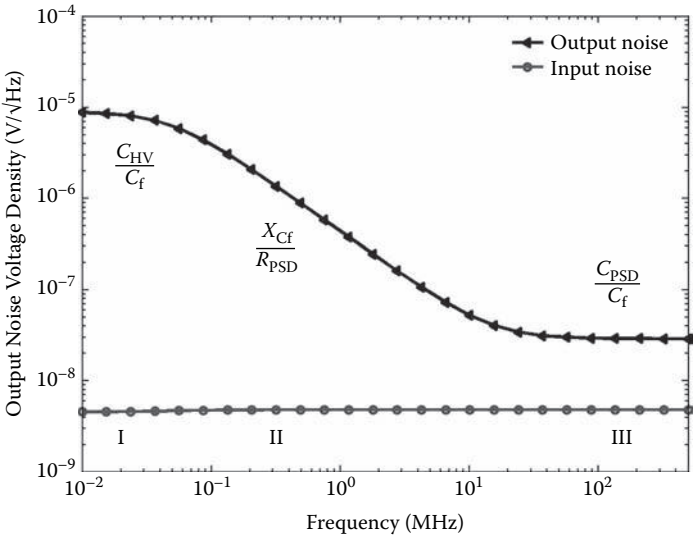



**FIGURE 10.26** Noise voltage density simulation of a coupled circuit (ideal OpAmp, 4.8 nV/√Hz equivalent input-noise voltage density, $R_f = 100$ MΩ, $C_f = 1.5$ pF, $C_{HV} = 3.3$ nF, $C_{PSD} = 15$ pF, noisy $R_{PSD} = 2.2$ kΩ).

$C_{HV}/C_f$ only. $C_{HV}$ may be reduced to reduce the low-frequency noise. However, the reduction of $C_{HV}$ also causes a lower charge equalization time and linearity errors in position determination [1]. An exact analytical/numerical determination of the optimum value depends very much on the specific detector system with its different design parameters, including the analog signal-conditioning circuit structure.

Input and output noise voltage densities calculated by a simulation using an ideal OpAmp model and an equivalent input noise voltage source supplying 4.8 nV/√Hz are shown in Figure 10.26. Cutoff frequencies and range borders may differ, depending on the values used in the actual device. Differences between measurement and simulation on the order of ±30% result from part value deviation ($C_f = 1.5 \pm 0.5$ pF).

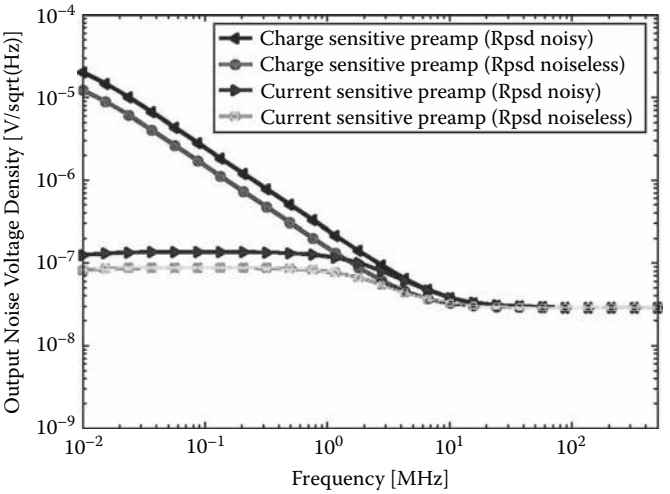



**FIGURE 10.27** Comparison of low-frequency noise using coupled CSA and ISA preamplifiers (ideal OpAmp, 4.8 nV/√Hz equivalent input-noise voltage density, $C_f$ = 1.5 pF, $C_{HV}$ = 15 nF, $C_{PSD}$ = 15 pF, $R_{PSD}$ = 4.8 kΩ; ISA: $R_f$ = 56 kΩ; CSA: $R_f$ = 100 MΩ).

It is important to distinguish the coupled noise from the noise described in the literature [3–6]. It is neither a simple input-referred white series noise nor an input-referred (series) 1/*f* noise. In the coupled case, the output noise is a gained white (series) voltage noise. Nevertheless, depending on the frequency, it shows a constant or 1/*f* behavior.

A noise analysis using a SPICE-compatible electric circuit simulator shows clearly that the thermal noise of the detector resistance—even though supplying about 6 nV/√Hz ($R_{PSD}$ = 2.2 kΩ) noise voltage density—can almost be neglected in the case of a PSD design using resistively coupled charge amplifiers having an equivalent input noise voltage density of 4.8 nV/√Hz [15].

A design using current-sensitive preamplifiers shows a certain advantage over the design using charge-sensitive preamplifiers (Figure 10.27). The low-frequency noise due to coupling is significantly reduced. Nevertheless, the mechanisms discussed previously must still be considered. The current-sensitive preamplifier design has been successfully tested in a position-sensitive neutron detector system [16].

### 10.4.3 Noise Correlation

In many publications, a very important assumption has been made, considering the correlation of noise at the two ends of the detector. The earliest reference inspiring confidence we found is a work by Radeka [3], which assumes that the noise voltages at the two ends of the detector are anticorrelated. We will understand that terminus to be "correlated with a noise-correlation coefficient of −1." Unfortunately, we did not find any proof for this claim in the literature. To subsequently prove this concept, we consider a symmetrical coupled circuit of two sensitive preamps (Figure 10.28).

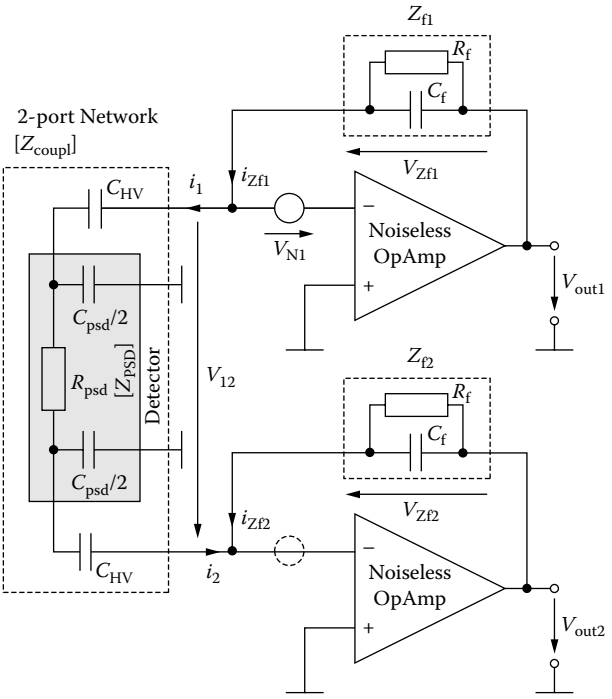



**FIGURE 10.28** Equivalent circuit for the noise-correlation analysis.

We will calculate the output voltages of the two preamps in the presence of one noise source. Due to the circuit symmetry, the contribution of the second noise source can be calculated accordingly. Here we sum up all coupling-related circuit elements (detector, coupling capacitors) into a two-port network ($Z_{coupl}$), an impedance matrix having four elements (input impedance $Z_{11}$, output impedance $Z_{22}$, forward transfer impedance $Z_{21}$, and reverse transfer impedance $Z_{12}$), and the feedback elements into the feedback impedance $Z_f$ ($Z_{f1} = Z_{f2} = Z_f$). Furthermore, the distributed resistances and capacitances to ground of the detector are taken into account using a first-order approach having two symmetrical capacitances ($C_{PSD}/2$) to ground. Actually, simulations have clearly shown that there is no difference in the noise correlation whether taking distributed components into account or not.

The upper output voltage is given by

$$v_{out_1} = v_{N_1} \cdot \left( 1 + \frac{X_{C_f} \parallel R_f}{X_{C_{HV}} + 2 \cdot X_{C_{PSD}} \parallel (R_{PSD} + 2 \cdot X_{C_{PSD}} \parallel X_{C_{HV}})} \right) \tag{10.40}$$

while the lower output voltage is given by

$$v_{out_2} = -v_{N_1} \cdot \frac{X_{C_f} \parallel R_f}{X_{C_{HV}} + 2 \cdot X_{C_{PSD}} \parallel (R_{PSD} + 2 \cdot X_{C_{PSD}} \parallel X_{C_{HV}})} \tag{10.41}$$

**TABLE 10.3**
**Frequency Ranges and Noise Correlation**

| Range | I | II | III |
|---|---|---|---|
| Frequency | <20 kHz | <5 MHz | >5 MHz |
| $X_{CHV}$ | $>R_{psd}$ | $<R_{psd}$ | $<R_{psd}$ |
| $X_{Cf}$ | $>R_{psd}$ | $>R_{psd}$ | $>R_{psd}$ |
| $X_{Cpsd}$ | $>R_{psd}$ | $>R_{psd}$ | $<R_{psd}$ |
| $V_{out1}$ | $\approx v_{N1}\frac{C_{HV}}{2C_f}$ | $\approx v_{N1}\frac{X_{Cf}}{R_{PSD}}$ | $\approx v_{N1}\left(1+\frac{C_{PSD}}{2C_f}\right)$ |
| $V_{out2}$ | $\approx -v_{N1}\frac{C_{HV}}{2C_f}$ | $\approx -v_{N1}\frac{X_{Cf}}{R_{PSD}}$ | $\approx -v_{N1}\frac{C_{PSD}}{2C_f}$ |
| Noise correlation coefficient | $\approx -1$ | $\approx -1$ | $\approx -\frac{C_{PSD}}{2C_f + C_{PSD}}$ $\approx -0.83$ |

The detailed analysis can be found in [15]. To draw conclusions from the output noise voltage calculation (Equations [10.40] and [10.41]), we need to consider three different frequency ranges (Table 10.3). Below 5 MHz (Ranges I, II), we really see a noise correlation with a correlation coefficient of −1, which means that the noise output voltages ($v_{out1}$, $v_{out2}$) swing into opposite directions. Above 5 MHz (Range III), the noise voltage contribution of one side to the other side is less for frequencies below 5 MHz and is frequency independent. Thus, in the upper frequency range, the noise is partly uncorrelated. The correlation depends on the detector capacitance $C_{PSD}$ and feedback capacitance $C_f$. The noise correlation can be improved with increasing detector capacitance or decreasing feedback capacitance.

### 10.4.4 Optimization

To investigate the trade-off between signal gain and noise gain for different device configurations, the classical CSA preamp and the ISA preamp using two different coupling resistances have been studied (Table 10.4).

**TABLE 10.4**
**Device Configuration of Preamplifier and Coupling Network**

| Value | Device 1 | Device 2 | Device 3 |
|---|---|---|---|
| $R_{psd}$ | 2.2 k | 4.8 k | 2.2 k |
| $C_{psd}$ | 15 p | 15 p | 15 p |
| $C_{HV}$ | 3.3 n | 15 n | 15 n |
| $R_f$ | 100 M | 56 k | 56 k |
| $C_f$ | 1.5 p | 1.5 p | 15 p |

The calculation (Figure 10.29) shows significant differences between the classical CSA preamp (Device 1) and an ISA preamp (Device 2). But it can be easily seen that in both cases the noise gain differs from the signal gain by more than three orders of magnitude, which proves the conclusions obtained previously.

The comparison of the ratio between signal gain and noise gain (Figure 10.30) suggests some advantages of the ISA preamp (Device 2) over a CSA preamp (Device 1). However, Device 3 proves that only the increase of the detector impedance is

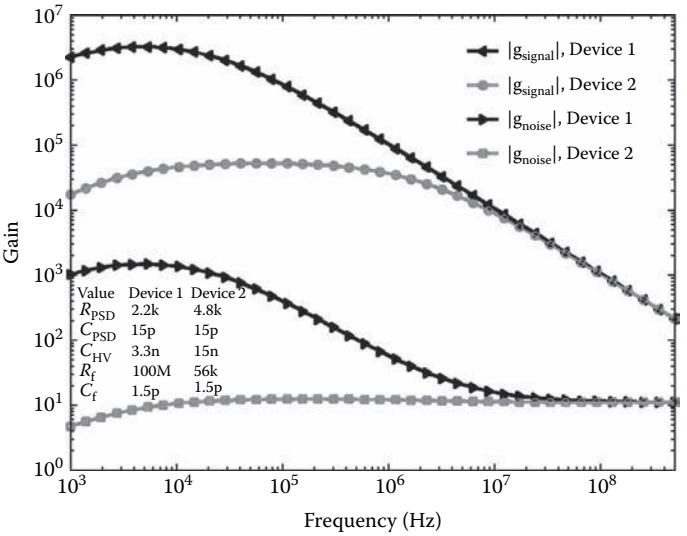



**FIGURE 10.29** Signal and noise gain calculation for a coupled circuit.

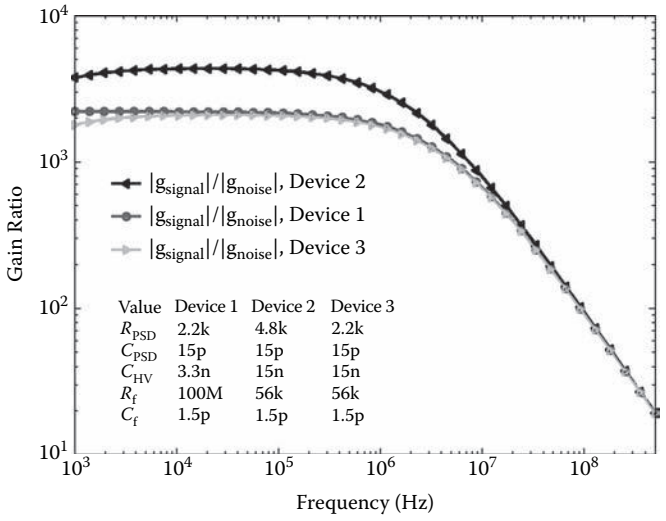



**FIGURE 10.30** Signal-to-noise gain ratio calculation for a coupled circuit.

responsible for a higher ratio of signal gain to noise gain. The configuration of the preamplifier either as a classical CSA (Device 1) or as an ISA (Devices 2 and 3) for frequencies of interest may be considered while optimizing position linearity. The noise gain of the ISA (Device 2) is around 10, which may suggest that the influence of the equivalent input noise source can be neglected and, due to the high signal gain, that only the thermal noise of the detector impedance is seen at the output. But this is not the case. As shown in Figure 10.30, the ratio of signal gain and noise gain stays almost constant. Furthermore, the detector impedance, mainly acting as a noise current source, supplies noise current that is about three orders of magnitude lower than noise voltage, which may partly equalize the effect of the high signal gain.

As it can be clearly seen that the detector resistance plays a major role for the system noise—not due to its own noise contribution but due to its coupling—there is a strong demand for its optimization. Different analytical optimization methodologies can be found in the literature. However, the availability of sophisticated design tools allows a deeper understanding of the parameters influencing system performance.

A complete signal-conditioning system design—consisting of two signal channels, each consisting of an ISA preamp and pulse shaping, with the coupling impedance including blocking capacitors and extension resistors (Figure 10.24) has been simulated using an electric circuit simulator. The signal-to-noise ratios for particle events at different relative positions and assuming different detector resistances considering the upper output $Q_1$ (close to relative position 0) are given in Figure 10.31. Higher detector resistances increase the charge division ratio. For an extremely low detector resistance (10 Ω), the SNR is a constant 43 dB for all relative positions. More realistic and therefore higher detector resistances increase the SNR for about 90% of the active detector region (i.e., below relative position 0.9). For 10% of the detector region (above relative positions of 0.9), the division ratio becomes so high that it will

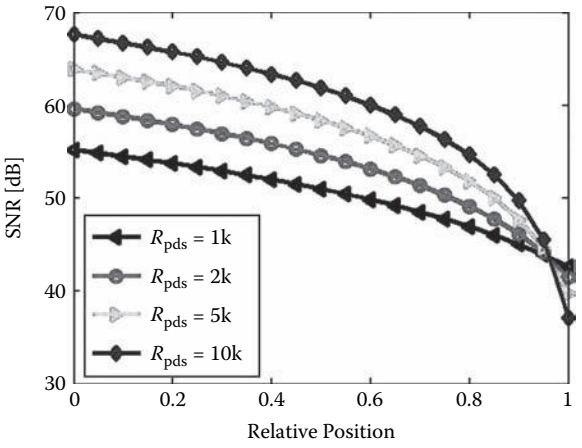



**FIGURE 10.31** Signal-to-noise ratio as a function of relative position of the particle event and detector resistance ($C_{HV}$ = 15 nF, $R_s$ = 270 Ω, $C_{PSD}$ = 15 pF; ISA: $C_f$ = 1.5 pF, $R_f$ = 56 kΩ; shaping: fifth-order Gaussian delay-time approximation, 1 μs FWHM; input signal: $I_{max}$ = 5 μA, $t$ = 500 ns, $q_{in}$ = 2.5 pC, single event).

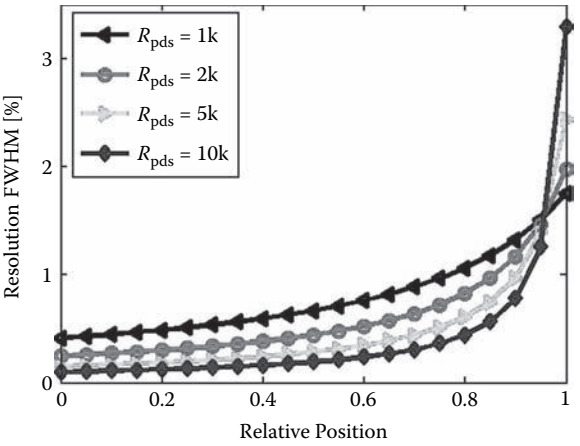



**FIGURE 10.32** Resolution as a function of relative position of the particle event and detector resistance ($C_{HV}$ = 15 nF, $R_s$ = 270 Ω, $C_{PSD}$ = 15 pF; ISA: $C_f$ = 1.5 pF, $R_f$ = 56 kΩ; shaping: fifth-order Gaussian delay-time approximation, 1 μs FWHM; input signal: $I_{max}$ = 5 μA, $t$ = 500 ns, $q_{in}$ = 2.5 pC, single event).

reduce the SNR significantly. For a realistic detector resistance of about 5 kΩ, the SNR ranges between 40 and 60 dB, depending on the position. The closer the event position to the output considered, the higher is the associated SNR.

Looking at the associated position resolution (Figure 10.32) reveals that, for 2.5 pC input charge and 5-kΩ detector resistance, the FWHM position resolution remains well below 1% over about 90% of the active detector region. A further increase of the input charge will improve the position resolution.

## 10.5 CONCLUSIONS

For excellent spatial and energy resolution, not only the detector itself, but also the analog signal-conditioning chain needs to be optimized. Even though the classical approach using a charge-sensitive preamplifier is well accepted and widely used, considering its structure, there is still room for improvement. The application of a newer approach using a current-sensitive preamplifier has the advantage of lower input impedance, pileup probability, and complexity as well as the opportunity for higher counting rates. The optimization of the position linearity of a coupled system is the most challenging design task. The application of an ISA, the choice of the right coupling impedance (i.e., blocking capacitors, series resistors, etc.), an appropriate shaping time, and a channel gain correction scheme ensure differential position nonlinearities ≤2%.

A pulse-shaping device can be used to form a desired output pulse shape. Although the well known $n$-time integration chain is easy understandable, the Gaussian delay-time approximation is found to be the preferable approach if a certain time domain error (i.e., over- and undershoots) needs to be specified. The approximation procedure and the device topology have been given to facilitate the design process of analog signal-conditioning systems.

An example design was given to demonstrate the design and performance of a current-sensitive signal-conditioning system considering ISA preamplifier stages and a Gaussian delay time approximation for the pulse-shaping device.

The use of the classical charge-sensitive approach in coupled systems (i.e., utilizing the charge division method) can cause excessive low-frequency noise. The reason is the difference of signal gain and noise gain causing a noise gain increase for frequencies below approximately 5 MHz, depending on the actual part values. This behavior can be found independently from the detector noise contribution. In that case, it can also be shown that the current noise contribution of the preamplifier (either CSA or ISA) is negligible. The noise gain of the white equivalent input (series) voltage noise in a coupled circuit is the reason for the strong increase of the output noise voltage. There are three significant ranges of noise gain to be considered. In the lower frequency range ($f < 20$ kHz), the noise gain is flat and determined by the ratio $C_{HV}/C_f$. In the mid-frequency range (20 kHz $< f <$5 MHz), the noise gain falls with $1/f$, determined by the ratio $X_{Cf}/R_{PSD}$. In the upper frequency range ($f > 5$ MHz), the noise gain is again flat and determined by the ratio $C_{PSD}/C_f$. A noise-correlation analysis has shown that the output noise voltages of two coupled preamplifiers are correlated up to 5 MHz by a noise correlation coefficient of −1. Above 5 MHz, the output noise is partly uncorrelated. The noise correlation can be improved by increasing the detector capacitance or decreasing the feedback capacitance.

The optimization process is a trade-off between different design parameters aiming at the highest position resolution and system linearity. As there are many parameters involved, optimization is conducted using sophisticated design tools. A channel gain correction scheme and signal-to-noise ratios in the range of 40 to 60 dB are needed for position resolutions well below 1% and integral linearity errors well below 0.1%.

## ACKNOWLEDGMENTS

I'd like to thank all persons at Helmholtz-Zentrum Berlin for their contributions to and support of this work. Especially I'd like to thank the head of the electronics department, Dr. F. Wulf, for his ongoing support and promotion for the realization of new ideas. I also kindly thank my colleague Dr. B. Namaschk for stimulating discussions and for his contributions to this publication.

## REFERENCES

1. Bönisch, S., B. Namaschk, and F. Wulf. 2007. Charge equalizing and error estimation in position sensitive neutron detectors. *Nuclear Instruments and Methods in Physics Research A* 570:133–139.
2. Radeka, V. 1968. State of the art of low noise amplifiers for semiconductor radiation detectors. *Proc. Intern. Nuclear Electronics* 1:46.
3. Radeka, V. 1974. Signal, noise and resolution in position-sensitive detectors. *IEEE Transactions on Nuclear Science* 21:51–64.
4. Radeka, V. 1988. Low-noise techniques in detectors. *Ann. Rev. Nucl. Part. Sci*. 38:217.

5. Bertuccio, G., and A. Pullia. 1993. A method for the determination of the noise parameters in preamplifying systems for semiconductor radiation detectors. *Rev. Sci. Instrum*. 64:3294–3298.
6. van Esch, P., T. Gahl, and B. Guerard. 2004. Design criteria for electronics for resistive charge division in thermal neutron detection. *Nuclear Instruments and Methods in Physics Research A* 526:493–500.
7. Xu, J., Y. Dai, and D. Abbott. 2000. A complete operational amplifier noise model: Analysis and measurement of correlation coefficient. *IEEE Trans. on Circuits and Systems*. 47:420–424.
8. Evangelisti, F., G. Orsi, and G. Ventura. 1973. A low noise, very low power charge sensitive amplifier for space applications. *Rev. Sci. Instrum*. 44:738–741.
9. Jones, A. O., and A. D. Smith. 1992. A high rate preamplifier for use with gas proportional counter. *Rev. Sci. Instrum*. 63:842–845.
10. ORTEC. Preamplifiers. Application note. Oak Ridge National Laboratory, Oak Ridge, TN.
11. Brisebois, G. 2005. Op Amp selection guide for optimum noise performance. Design Note 355. Milpitas, CA: Linear Technology Corp.
12. Texas Instruments. 2006. 1.6GHz, low-noise, FET-input operational amplifier. Datasheet OPA657. Dallas: Texas Instruments.
13. Spieler, H. 2005. *Semiconductor detector systems*. Oxford, U.K.: Oxford University Press.
14. Lutz, G. 1999. *Semiconductor radiation detectors*. Berlin: Springer-Verlag.
15. Bönisch, S., B. Namaschk, and F. Wulf. 2008. Low-frequency noise of resistively coupled charge amplifiers. *IEEE Transactions on Nuclear Science* 55:2315–2322.
16. Bönisch, S., B. Namaschk, and F. Wulf. 2008. PSD-readout electronics development for the new EXED experiment at HZB. Paper presented at IEEE Nuclear Science Symposium, Dresden.
17. Knoll, G. F. 2000. *Radiation detection and measurement*. New York: Wiley.
18. Borderie, B., et al. 2004. Charge and current-sensitive preamplifiers for pulse shape discrimination techniques with silicon detectors. *Nuclear Instruments and Methods in Physics Research A* 531:607–615.
19. Moschytz, G. S., and P. Horn. 1981. *Active filter design handbook*. New York: Wiley.
20. Fritzsche, G. 1962. *Entwurf linearer Schaltungen*. Berlin: Verlag Technik.
21. Solberg, K., et al. 2007. Square single-wire detectors for neutron diffraction studies. *Nuclear Instruments and Methods in Physics Research A* 580:1110–1114.

# 11 Analog-to-Digital Converters for Radiation Detection Electronics

*Rafal Dlugosz and Krzysztof Iniewski*

## CONTENTS



## 11.1 INTRODUCTION

The ability to peer into the human body is an essential diagnostic tool in medicine. Many of the imaging modalities, including computed tomography (CT) and SPECT/PET (single photon emission computed tomography/positron emission tomography) nuclear medicine techniques, are based on X-ray or γ-ray transmission and detection. In addition, radiation detection is used in numerous other fields, including high-energy physics, security, luggage scanning, and space applications. Despite their different principles of operation, there are numerous commonalities in the processing of signals received by these imaging detectors: signal amplification, filtering, multiplexing, and analog-to-digital conversion (ADC). The last step, A/D conversion, is essential, as gathered information is typically processed using digital signal processing algorithms.

The X-ray detection systems can be separated into direct- and indirect-conversion architectures. In the first case, a scintillator material is used to convert the photons to light that is then subsequently converted to an electrical signal using photodiodes, charge-coupled devices (CCDs), or CMOS (complementary metal-oxide semiconductor) imaging sensors. On the other hand, the direct-conversion concept, which is currently gaining popularity, converts photons directly into an electrical charge, which allows for a better image quality.

The block diagrams of two possible direct conversion schemes are shown in Figure 11.1. In the first approach, shown in Figure 11.1(a), the analog output signals of the charge-to-voltage (C/V) converters are serialized in the asynchronous multiplexer (MUX), stored in the first-in-first-out (FIFO) block, and then converted to a digital signal using a single high-performance ADC that usually is an external block in such systems. In the second approach, shown in Figure 11.1(b), each channel in the C/V or in the charge-to-current (C/I) converter is connected to a separate low-power and low-chip area ADC. Particular ADCs are active only when a given channel detects a photon. In this case, the analog FIFO block is not required, which simplifies the control circuitry. This additionally minimizes the number of copy operations at the analog side, which potentially increases the overall system performance.

The frequency content of different imaging signals covers different portions of the spectrum and signal bandwidths (BW), and can range from near DC to several kilohertz (kHz) and possibly up to a few megahertz (MHz). Some measured signals have higher amplitudes than others, and they range from a few microvolts (μV) to tenths of millivolts (mV). Both low-voltage operation and low-power dissipation are of great importance for these types of applications, because hundreds or thousands of imaging sensors have to be processed simultaneously. Low-voltage operation is demanded because it is desirable to use as few batteries as possible for size and weight considerations. Low current consumption is necessary to ensure no increase in sensor operating temperature and a reasonable battery lifetime for portable equipment.

Typical data acquisition systems have several performance limitations. They are normally based on ADCs with resolutions of between 8 and 14 bits, meaning that there is a relatively high dynamic range to deal with simultaneously strong noise disturbance signals (such as the 50/60-Hz power noise signal) of comparable magnitude. As CMOS technologies continue to evolve toward smaller geometries, new design techniques are being developed to improve the power efficiency of these ADCs while possibly reducing the silicon die area.

This chapter is devoted to the A/D converters used in medical imaging systems. We will briefly discuss ADC configurations that are suitable considering basic properties of flash, pipeline, and successive approximation (SAR) converters. A basic review of existing architectures indicates that the most power efficient converters are built using the SAR principle. For this reason, most of the material that follows will be devoted to SAR analog-to-digital converters. A review of existing SAR ADC solutions, new design techniques, and design examples will be presented and discussed. For other types of ADC architectures, readers can consult several books dedicated entirely to this topic [1, 28].

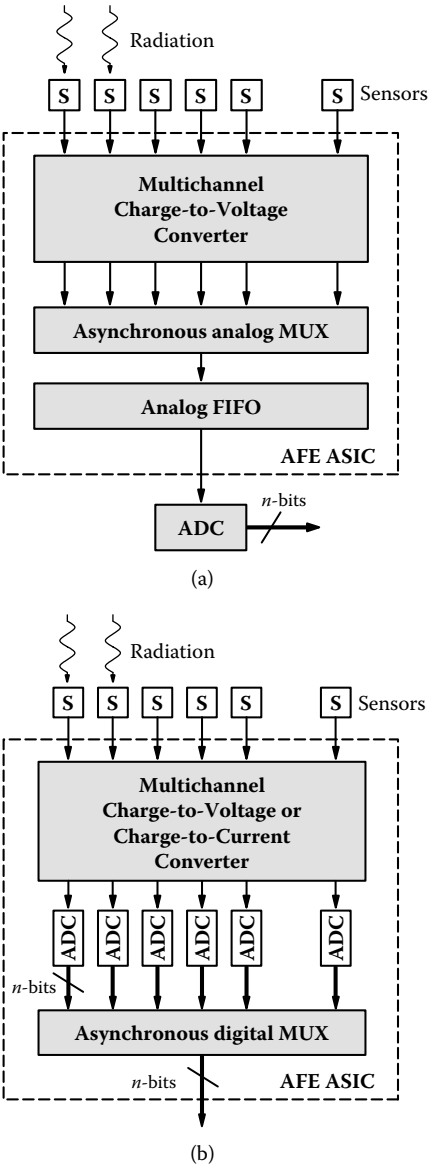



**FIGURE 11.1** Direct conversion architecture: (a) with analog FIFO structure and an external ADC; (b) without the analog FIFO block and a number of low-power and low-chip-area ADCs working in parallel, integrated with other blocks in the charge-to-voltage ASIC.

## 11.2 GENERAL CLASSIFICATION OF ADCS

Analog-to-digital converters can be classified as Nyquist rate ADCs and oversampled ADCs (sigma–delta). Sigma–delta ADCs are preferred for the highest bit resolutions; however, they demand very fast oversampling clocks, which makes them inherently

low-speed structures. Their clock generator has to be very clean, as the clock jitter has a direct influence on the obtained signal-to-noise ratio (SNR).

On the other hand, the Nyquist rate converters are typically used in applications that require lower data resolution and higher data rates. Nyquist rate converters usually have one of the following four architectures [1, 28]:

1. Flash
2. Pipeline
3. Folding–interpolating
4. Algorithmic and successive approximation register (SAR)

Taking into consideration such features as attainable data conversion rates, output data resolution, and power dissipation, particular converter architectures are preferred in different types of applications. SAR converters with only one comparator are simple structures with relatively low power dissipation, but as only one bit is calculated per one clock cycle, they are also relatively slow. In the applications where a high conversion rate is required, SAR ADCs put high demands on the comparator, which must be at least $n$ times faster than comparators in flash converters for the same data rate, where $n$ is the resolution of the output data. SAR converters require a very good calibration of the digital-to-analog converter (DAC), since the DAC accuracy has a direct influence on linearity of the input–output characteristic. In SAR converters, all bits are calculated under equal conditions using the same comparator. Hence, the offset of the comparator does not have influence on the linearity of the input–output characteristic. However, the offset can shift the characteristic by a constant DC value. To achieve a short conversion time, several SAR ADCs can be time interleaved; however, in this case, power budget and nonuniformities of ADC characteristics become more critical.

In flash ADCs, all bits are calculated in parallel, which makes this type of conversion potentially the fastest of all approaches. The state-of-the-art converters of this type implemented in a standard CMOS process operate with sampling rates of up to several gigahertz (GHz) [62–64], while the rates realized in SiGe technology can be up to several dozen gigahertz [65]. In this architecture, offset compensation is very important, since different offsets of comparators operating in a thermometer-type conversion have a direct influence on the differential nonlinearity (DNL) of the input–output characteristics. The price for a high resolution and a fast conversion is very high power dissipation, as the number of comparators is equal to $2^{n-1}$. For this reason, flash converters are used in high data rate applications that accept low resolutions, typically in the range of 3 to 6 bits, and where power dissipation is of a secondary importance. Their application in medical imaging is virtually nonexistent, besides PET systems, where they are typically used as a low-resolution component in time-to-digital schemes [67, 68].

Taking power dissipation and data rate criteria into account, SAR and flash ADCs can be viewed as two extreme cases [57], while other ADC architectures such as pipelined and folding–interpolating can be seen as the intermediate solutions. The last two types of converters usually are designed for higher resolutions (14–20 bits) [61], but present a trade-off of the number of comparators, the conversion rate, and

the power dissipation. For example, low-resolution flash ADCs are used in particular $k$-bit stages of pipelined ADCs. This minimizes the total number of comparators when compared with pure flash solutions, but conversion of each data point requires several clock cycles. In general, pipelined and folded–interpolated architectures have much more complex structure than SAR converters, since they require a more complex control block [61].

Figure 11.2 shows a comparison of existing converter solutions in terms of data rate versus dissipated power for selected converters of different types [2–27, 31–60, 62–64]. All presented results have been normalized to 8 bits of resolution and with respect to $V_{DD}{}^2$, where $V_{DD}$ is the power supply level. As can be observed, the dissipated power correlates well with the sampling frequency, as expected.

Each group of converters is clustered in one area of the plot, although these areas slightly overlap. The placement of particular groups corresponds to the previous description. The most flexible architecture in terms of the power dissipation and the sampling frequency is the first group. Algorithmic-SAR converters usually are designed for low sampling frequencies and low power dissipation, although time-interleaved architectures have recently been gaining popularity [11, 44–48]. These converters use relatively slow SAR blocks working in parallel. Due to the low number of comparators that is equal to data resolution, $n$, instead of $2^n$ as in flash

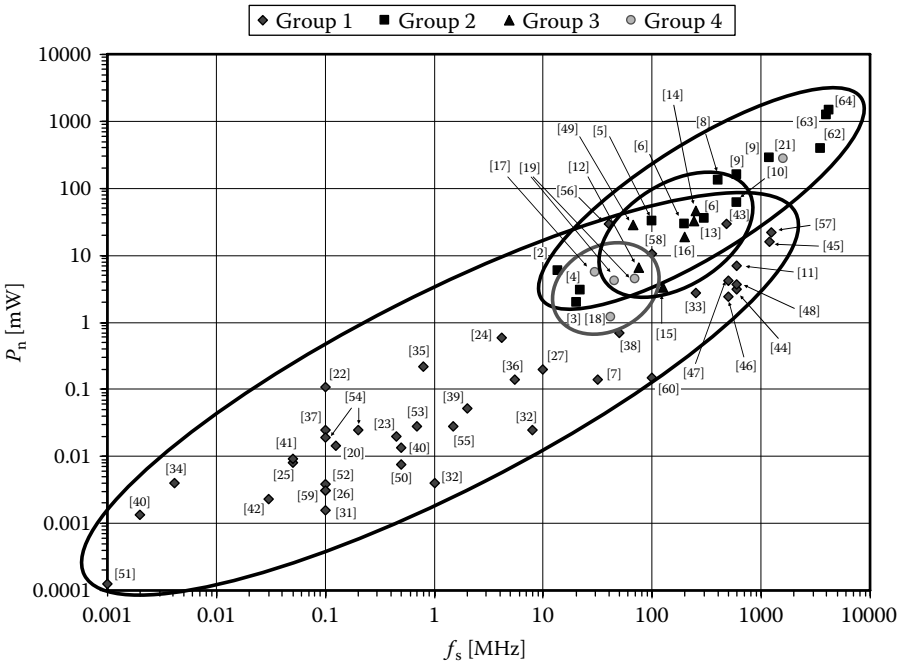



**FIGURE 11.2** Performance comparison of four groups of ADCs: Group 1, algorithmic and SAR; Group 2, flash; Group 3, pipelined; Group 4, folding and interpolating. Normalized power with respect to $V_{DD}$ = 1 V. (Based on Dlugosz and Iniewski [32]. © 2007 by Hindavi Publishing. With permission.)

ADCs, these converters can effectively compete in terms of sampling frequencies, simultaneously dissipating much less power. Comparing, for example, converters [9] with [45] and [8] with [46], we see that power dissipation differs by as much as one order of magnitude for similar sampling frequencies.

## 11.3 SAR PRINCIPLE OF OPERATION

A typical classic SAR converter contains four main components, as shown in Figure 11.3, i.e., the sample-and-hold element (S&H) that is used to acquire the input signal, a comparator, a digital-to-analog converter (DAC), and the control-logic circuitry that controls the calculation scheme.

In SAR ADC, a binary search algorithm is used in the conversion of the analog signal into its digital representation. This algorithm consists of several steps. Initially after resetting the logic circuitry and DAC circuit, the most significant bit (MSB) is set to digital 1. The DAC block supplies the analog equivalent signal of this code into the comparator circuit for comparison with the sampled input signal that is stored in the S&H circuit. Depending on which signal is larger (input or reference), the comparator either resets or maintains this bit. In the next step, the second bit is set up to a digital 1, and the DAC block supplies a new value for the reference signal. This binary search is performed for all bits until the least significant bit (LSB) is tested.

SAR converters usually are implemented using the voltage-mode charge-redistribution architecture shown in Figure 11.4. The most characteristic block in this approach is a charge-scaling DAC, which consists of an array of individually switched binary-weighted capacitors. These capacitors also serve as an S&H memory element. A whole conversion cycle performed in this converter consists of three initial steps followed by a conversion algorithm that is then performed in $n$ phases.

Each conversion cycle starts with discharging of the capacitor array. A good technique is to discharge capacitors to the offset voltage of the comparator instead of the ground, as this enables automatic offset cancellation. In the next step, free terminals of all capacitors are switched to the input voltage $V_{in}$, while the common terminal is switched to the ground (acq = 1 and sar = 0). The amount of charge that is switched to the particular capacitors is equal to their capacitance times the input voltage. In the third step, the common terminal is disconnected from the ground and remains connected to the comparator's negative input, while all free terminals are connected to the ground ($b_i = 0$, for $i = 1, \ldots, n$). As a result, the comparator input voltage

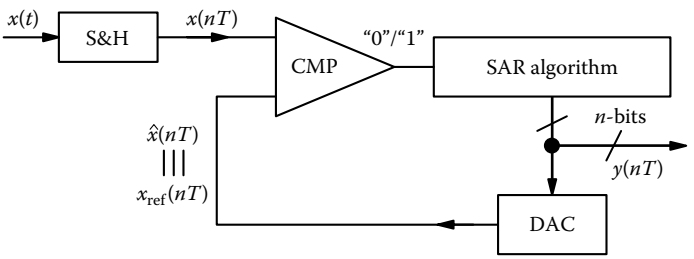



**FIGURE 11.3** A schematic block diagram of the successive approximation ADC.

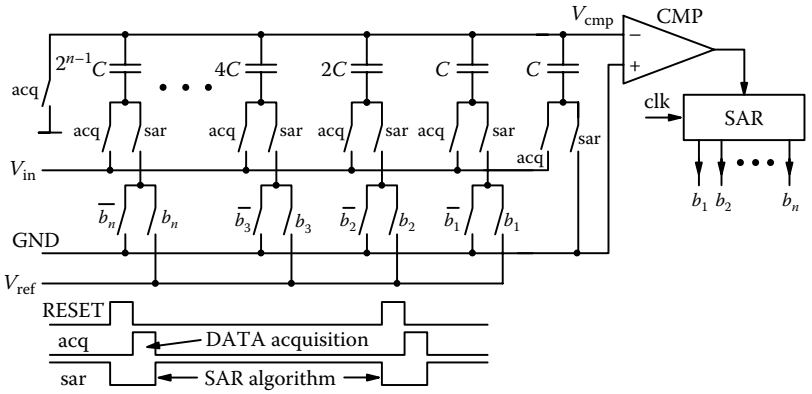



**FIGURE 11.4** Schematic diagram of the charge-redistribution SAR ADC architecture.

equals the negated value of the input signal. Finally, the SAR conversion algorithm starts in the fourth step.

In the following $n$ clock cycles, particular binary-weighted capacitors, from MSB to LSB, are switched to the reference voltage $V_{ref}$ that corresponds to the full-scale voltage ($V_{FS}$) range of the ADC. First, the MSB capacitor is switched to $V_{ref}$ ($b_n = 1$). As a result, the comparator's negative input voltage becomes $-V_{in} + V_{ref}/2$. Depending on the comparator's output, the first capacitor remains connected to $V_{ref}$ voltage or is switched again to ground. In the next clock cycle, the second capacitor in the array is connected to $V_{ref}$, and the comparator's input becomes equal to $-V_{in} + V_{ref}/2 + V_{ref}/4$ or to $-V_{in} + V_{ref}/4$, depending on the previous comparison result. This sequence is repeated for all bits, so the comparator-negated input voltage can be expressed as follows:

$$V_{cmp} = -V_{in} + V_{ref} \sum_{i=1}^{n} \frac{b_i}{2^{n-i+1}} \tag{11.1}$$

SAR converters can also be implemented using current-mode circuits. In this case, the DAC block is realized as a multi-output current mirror, with particular output transistors being binary weighted. In current-mode SAR converters, the initial phase is shorter compared to charge-redistribution solutions. In this case, sampling of the input signal in the S&H element can be performed in parallel with resetting the logic circuitry. The conversion algorithm is similar to that performed in the voltage-mode approach. Particular bits set up the value of the reference current, which is compared with the input signal in the current-mode comparator.

Current-mode SAR converters have several advantages over charge-redistribution architectures. A shorter initial phase is one such advantage. Another one is the very small amount of energy consumed during the calculation of a single bit. Currents that flow in the current-mode DAC must charge only parasitic capacitances, in particular transistors, which may be even several orders of magnitude smaller than capacitances in charge-redistribution converters. As a result, current-

mode converters are more power efficient and potentially faster. An additional advantage is very small chip area, since the DAC block in this case does not contain large capacitors.

The disadvantage of current-mode SAR converters is lower resolution than in charge-redistribution solutions. Since the DAC block in this case is realized as the current mirror, the transistor-matching problem occurs. This problem, which has an influence on the linearity of the input–output characteristics, becomes especially important for small currents and small power supplies when transistors work in a weak inversion region. Since the DAC block in charge-redistribution converters is implemented using capacitors whose matching does not depend on the value of supply voltage, this problem is less insignificant in this case.

## 11.4 SAR VERSUS FLASH ADC

An exact comparison of SAR and flash converters has been performed by Ginsburg and Chandrakasan [46] on the basis of developed analytical energy models. These models enable the selection of a given architecture for a given application in cases where power dissipation is the main criterion. In flash architectures, energy is consumed mostly by comparators, while in SAR converters, all building blocks described earlier must be considered in the energy model. In SAR ADCs, particular energy components have different distributions over the converter's resolution, as shown in Figure 11.5.

In a comparator with a two-stage preamplifier, energy varies almost linearly with the converter's resolution. The comparator must make the proper decision even when both input voltages differ insignificantly. When the converter's resolution increases,

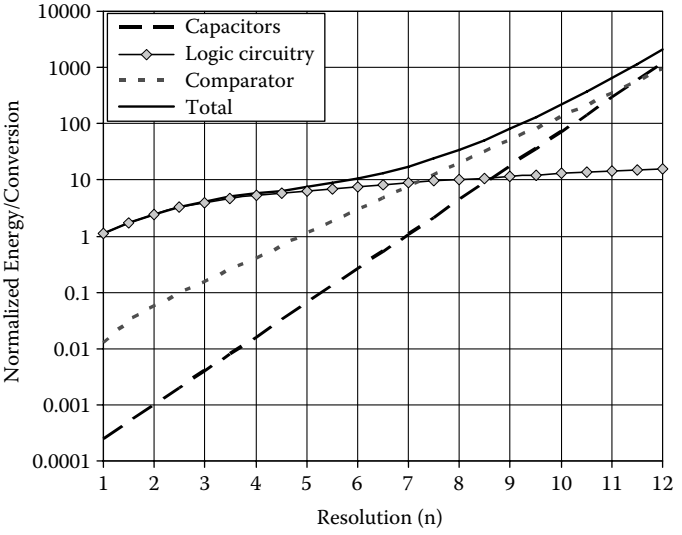



**FIGURE 11.5** Energy versus resolution in SAR converters for particular building blocks such as comparator, DAC, and control logic. (From Ginsburg and Chandrakasan [46]. © 2007 by IEEE. With permission.)

then the minimal difference between inputs decreases, and preamplifiers must have higher gains to ensure the appropriate settling time of the comparator.

Energy consumed in the logic circuitry increases only moderately with the resolution, as each algorithm step requires almost the same control block. The moderate increase that is visible in Figure 11.5 is due to the different dimensions of switches in particular stages. In cases of larger capacitors, the on-resistance of switches must be smaller to keep the settling time constant in all stages. Energy that is consumed by the capacitor array increases exponentially with the converter's resolution, which is due to the exponential growth of capacitance in the following DAC stages.

Models presented by Ginsburg and Chandrakasan [46] take into consideration the influence of the full-scale input voltage ($V_{FS}$) on the energy of the converter. When the value of $V_{FS}$ increases in a SAR converter, then a bigger charge is fed to the capacitors in each conversion cycle. On the other hand, a higher value of $V_{FS}$ minimizes the gain requirements of preamplifiers in the comparator, resulting in a lower level of energy consumed by this element. The decrease of energy consumption in the comparator compensates the increase of energy in the capacitors. As a result, in SAR converters, energy varies only moderately with $V_{FS}$. In the case of flash converters, where energy is consumed mostly by comparators, higher values of $V_{FS}$ enable significant improvement.

Published models show that each converter is suitable for different values of parameters. For example, when the resolution is higher than 4 bits, then for a $V_{FS}$ that is equal to 300 mV, SAR converters exhibit better efficiency than flash converters, as shown in Figure 11.6(a) [46]. For a given resolution when $V_{FS}$ is higher than some threshold value, then flash architecture attains better power efficiency. This is illustrated in Figure 11.6(b) for an example resolution of 5 bits.

## 11.5 STATE OF THE ART IN SAR ADC DESIGN

SAR converters are usually designed for resolutions between 8 and 10 bits, although higher resolutions up to 14 bits are also reported [37, 43]. Higher resolutions require special techniques that enable a proper calibration of DAC to ensure a high linearity of the input–output characteristic. For example, in the converter described by Hesener et al. [43], the capacitor array is realized as a cluster of small unity capacitors with redundancy in the number of these unity capacitors. Configuration of the array is controlled by an additional logic circuit, which at each algorithm stage calculates the required capacitors values and programs the capacitor array. This mechanism allows for the correction of capacitor mismatch, but makes the structure more complex and increases both the chip area and power dissipation that is, in this case, the highest of all SAR converters presented in Figure 11.2.

Most of the reported SAR ADCs have been designed as voltage-mode architectures. In the classic charge-redistribution scheme, larger resolutions are difficult to attain, as capacitance values for each additional bit must be doubled. This doubles also the power dissipated in the capacitor array during data acquisition. Minimizing the value of the unity capacitor brings no effect, since it worsens matching between capacitors.

An example implementation of classic charge-redistribution converter has been described by Sauerbrey, Schmitt-Landsiedel, and Thewes [34]. This 9-bit converter

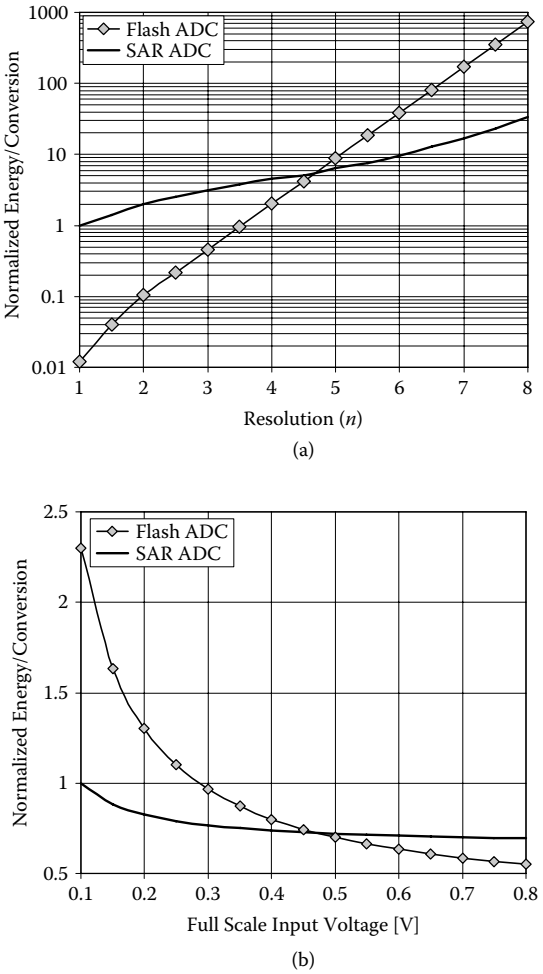



**FIGURE 11.6** A comparison of ADC for SAR and flash converters: (a) energy versus resolution for $V_{FS} = 300$ mV, (b) energy versus full-scale input voltage for resolution of 5 bits. (From Ginsburg and Chandrakasan [46]. © 2007 by IEEE. With permission.)

for the sampling frequency of 150 kHz dissipates power of 30 μW from a 1-V power supply, and for 4.1 kHz dissipates power of 0.85 μW from 0.5 V. In this case, the unity capacitor has capacitance that is equal to 20 fF, while the biggest capacitor has capacitance of 5 pF, which makes the overall capacitance in the circuit larger than 10 pF. For 12 bits of resolution, the overall capacitance in the circuit would be larger than 80 pF. This would limit the possible sampling frequency by several times for a given value of the power dissipation. An additional problem related to the capacitor array is the large chip area of the array that, in this case, is equal to about 70% of the overall chip area. All these problems show practical limitations of the classic charge-redistribution architecture. To overcome these problems, various optimization techniques are used.

### 11.5.1 Optimization of the Capacitor Array

One of the possibilities to minimize power dissipation is minimizing the value of the unity capacitor, but this technique increases the matching error. As a result, it creates a trade-off between power dissipation and the linearity of the converter. This problem can be solved in several ways. For example, in the 8-bit converter reported by Scott, Boser, and Pister [26], a careful design of the capacitor array enabled a high linearity of DAC for a small unity capacitor with a capacitance of only 12 fF and an overall capacitance of 3 pF. This converter has also been designed using the classic charge-redistribution architecture, but due to several optimization techniques, power dissipation (3.1 μW) is in this case almost one order of magnitude smaller than in the case of a converter reported by Sauerbrey, Schmitt-Landsiedel, and Thewes [34] for the same sampling frequency of 100 kHz. Poly–poly capacitors have been used with top plates used as common plates. This technique allows for the minimization of the parasitic capacitance at the comparator input. The dummy capacitors have been placed around the capacitor array, and they improve matching between capacitors and the linearity of the DAC. The other interesting technique used in this converter is placing a grounded metal shield over the poly–poly capacitors and then routing the capacitors above this shield. This technique makes the capacitors array insensitive to routing parasitics, which protects the array against systematic matching errors. The resultant DNL and INL (integral nonlinearity) parameters are kept on a very low level of ±0.25 LSB.

Another optimization method of the capacitor array that has gained some popularity in recent years takes advantage of the split-array technique [40, 52, 66]. In this technique, instead of using the binary-weighted capacitor array—as in the converters reported by Scott, Boser, and Pister [26] and Sauerbrey, Schmitt-Landsiedel, and Thewes [34]—an additional attenuation capacitor, $C_{att}$, divides the capacitor array into two subarrays responsible for calculation of LSBs and MSBs, respectively, as shown in Figure 11.7. When capacitance of the attenuation capacitor has a value that satisfies the following equation:

$$C_{att} = C\frac{\sum(\text{LSB capacitors})}{\sum(\text{MSB capacitors})} \tag{11.2}$$

then the resultant values of LSB capacitances that the comparator sees at point A are smaller by a factor of $2^i$, where $i$ is the number of LSB capacitors (excluding a dummy capacitor $C_d$).

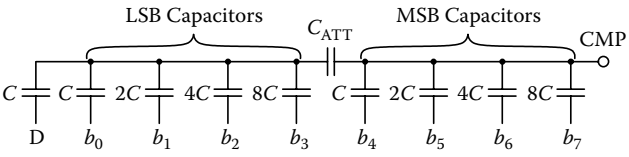



**FIGURE 11.7** A split-capacitor array that enables reduction of the ratio between the largest and smallest capacitors in a DAC block. (From Abdelhalim, MacEachern, and Mahmoud [40]. © 2007 by IEEE. With permission.)

This technique allows a significant reduction of the spread between the smallest and the largest capacitors in the array, maintaining capacitors in MSBs that are relatively small. In the converter described by Abdelhalim, MacEachern, and Mahmoud [40], this allowed an increase of the unity capacitor, which improved matching. The unity capacitor has, in this case, a value of 211 fF, whereas the largest capacitor in the circuit is equal to only 1.56 pF (compared to 20 fF and 5 pF, respectively, in the capacitor described by Sauerbrey, Schmitt-Landsiedel, and Thewes [34]). The value of the attenuation capacitor $C_{att}$ is equal to 225 fF. The smallest capacitor in the LSB area has a value of only 13 fF. These values are similar to those in the converter presented by Scott, Boser, and Pister [26], and the resultant sampling frequencies and power dissipation are also similar. For example, this converter dissipates power of 8.5 μW for 500 kS/s and less than 1 μW for 60 kS/s. In the converter described by Agnes et al. [52], the unity capacitor is 120 fF, while the largest one is equal to 3.8 pF.

### 11.5.2 Optimization of the Comparator

Improvement in SAR ADC's performance is also possible through the optimization of the comparator. The energy model introduced by Ginsburg and Chandrakasan [46] shows that offset of the comparator has a direct influence on the effective number of bits (ENOB), i.e., on the output data resolution. The optimization techniques, in this case, rely on minimizing the offset or increasing the gain of preamplifiers in comparators, which enables offset compensation. In the converter reported by Verma and Chandrakasan [31], several optimization techniques have been used. One of them introduces an offset-compensating regenerative latch into the comparator, which enables a decreased gain of the preamplifier, improves the power-delay product (PDP), and decreases the propagation delay in the comparator. The settling time has been additionally improved by introducing the self-timing technique. As a result, resolution of 12 bits is possible in this converter for sampling frequency of 100 kHz and power dissipation on the level of 25 μW. For comparison, in the converter reported by Sauerbrey, Schmitt-Landsiedel, and Thewes [34], an ENOB of 9 bits has been achieved for 30 μW of power.

The other offset correction technique has been introduced in the converter design reported by Abdelhalim, MacEachern, and Mahmoud [40]. The comparator scheme is shown in Figure 11.8. In this block, during the sampling phase, both comparators' terminals are switched to $V_{SS}$, the switch controlled by the RST signal is closed, and the gates of transistors M9 and M10 are on the common potential, which is equal to $V_{DD}/2$ plus an offset voltage. This value is stored across the capacitor $C_{off}$. During the conversion phase, the switch is opened and the output signal from the first stage is compared with the signal that is stored across capacitor $C_{off}$. This technique allows for a reduction of the offset of more than 60%. It is worth noting that this converter can operate in a relatively wide range of the supply voltage between 0.36 and 0.8 V. When operating with the small value of supply, the converter dissipates only 230 nW for the sampling frequency equal to 20 kHz. This is the smallest power dissipation reported so far. This result is much better than in the converter reported by Sauerbrey, Schmitt-Landsiedel, and Thewes [34], where for the supply voltage of 0.4 V and the power dissipation equal to 280 nW, the sampling frequency is equal to only 600 Hz.

Another solution, which can be referred to as nonconventional, has been proposed by Agnes et al. [52]. This SAR ADC uses a time-domain comparator that allows for realization of a power-efficient converter. The comparator is shown in Figure 11.9. It consists of two voltage-to-time (V2T) converters and a logic circuit. Transistors

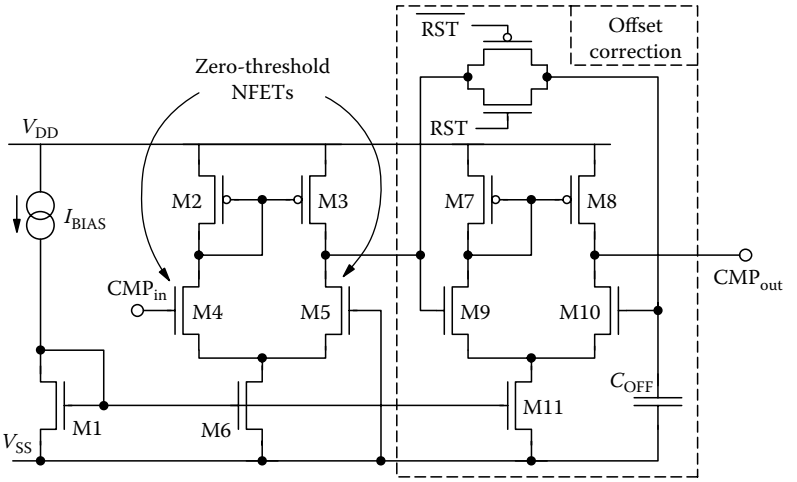



**FIGURE 11.8** A comparator with offset correction used in a converter. (From Abdelhalim, MacEachern, and Mahmoud [40]. © 2007 by IEEE. With permission.)

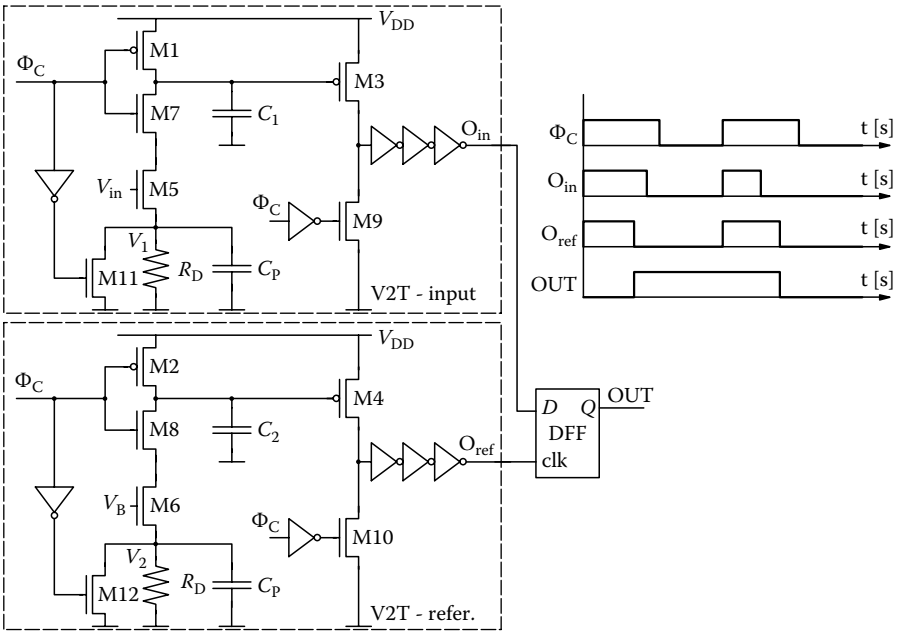



**FIGURE 11.9** Time-domain comparator. (From Agnes et al. [52]. © 2008 by IEEE. With permission.)

M1 and M2 in V2T blocks charge the nominally equal capacitors C1 and C2. When the voltages across the C1 and C2 capacitors cross the threshold voltages of M3 and M4 transistors, the chains of inverters switch, producing two logical 1 signals with different delays that depend on values of the input voltages. A DFF (D-type flip-flop) circuit at the output, which serves as a time comparator, determines which V2T is faster, i.e., which logical 1 occurs first. The comparator operates with less than 1 μW of power at 1-V supply and enables sampling frequency of 1.4 MHz with 0.2-mV sensitivity. Comparing this ADC with a solution reported by Scott, Boser, and Pister [26], the sampling frequency is, in this case, more than one order of magnitude higher for a similar power dissipation of 3.8 μW. The resolution of 10 bits is also higher in this case. The possible offset in the comparator is, in this case, due to a mismatch between transistors used in both V2T blocks. In the case when both input voltages are equal, the mismatch causes the currents that charge the capacitors C1 and C2 to differ slightly. The offset can be corrected by adjustment of the bias voltage $V_B$.

## 11.6 DESIGN EXAMPLE OF 8-BIT CURRENT-MODE SAR

Although most algorithmic-SAR converters have been designed using the charge-redistribution approach described earlier, in the case of low-power, low-chip area and low-to-moderate sampling frequency applications, current-mode algorithmic or SAR converters are gaining popularity [20, 23, 27, 29, 32, 60]. Considering direct-conversion architecture with multiple ADCs, shown in Figure 11.1(b), a chip area of a single ADC is the parameter as important as the energy consumption. A small area can be achieved only in current-mode SAR ADCs, which do not require large capacitors in DAC. As a result, the die size of current-mode ADCs can be 20–50 times smaller than that of voltage-mode converters. For example, the converter described by Agnes et al. [52] occupies an area of 0.24 mm$^2$, while an example current-mode converter reported by Yang and Van der Spiegel [60] occupies an area of only 0.005 mm$^2$, and another converter described by Dlugosz and Iniewski [32] occupies an area of 0.009 mm$^2$.

An example is the current-mode SAR converter shown in Figure 11.10 [32]. It consists of the analog circuitry responsible for signal conversion and the control-logic block. In this case, both analog and digital parts are powered from separate power supplies to enable power-down modes and to reduce digital-to-analog noise cross talk.

This converter derives its essence from the typical integrating current comparator realized on the CMOS inverters. This circuit converts an input current $I_{in}$, which is obtained by using a current-mode sample and hold (S&H) element. The input current is compared in each approximation step with a reference current $I_{ref}$, which is the output of the current-mode DAC and can be expressed as

$$I_{ref} = (b_n \cdot 2^{n-1} + b_{n-1} \cdot 2^{n-2} + \ldots + b_2 \cdot 2 + b_1) \cdot I_{tr} \tag{11.3}$$

where $I_{tr}$ is a biasing current. The differential current, $I_c$, between $I_{in}$ and $I_{ref}$ ($I_c = I_{in} - I_{ref}$) changes the voltage across the input gate-to-source capacitances of the first inverter

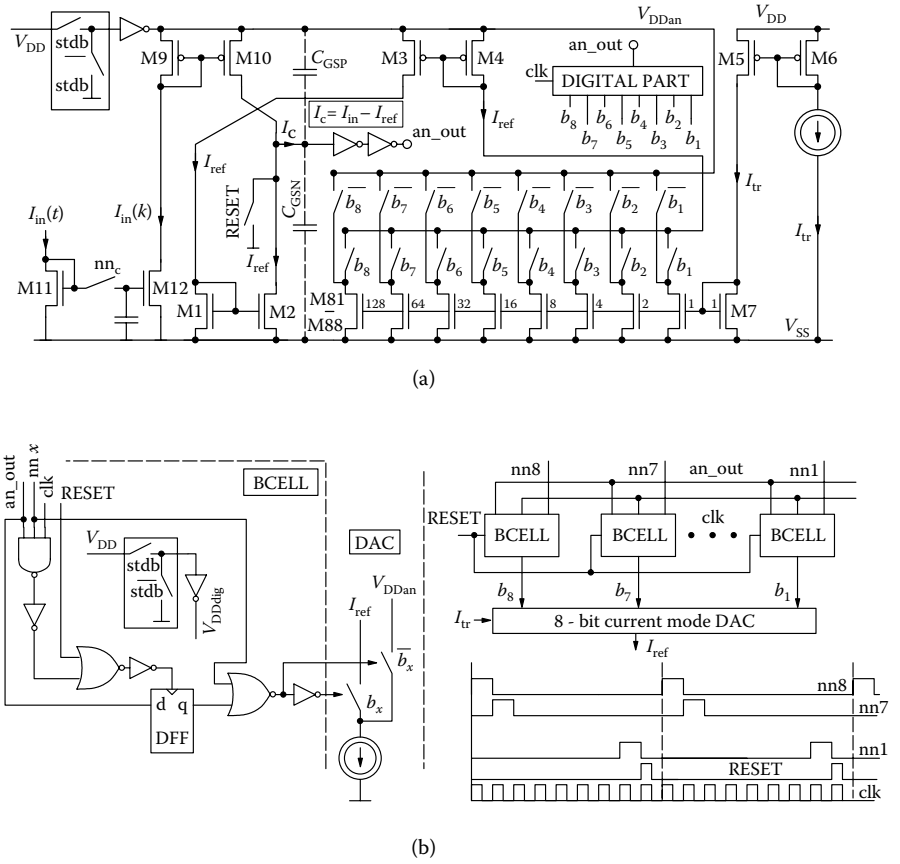



**FIGURE 11.10** A block diagram of an example current-mode SAR ADC with an implemented standby mode: (a) analog part and (b) digital part. (From Dlugosz and Iniewski [32]. © 2007 by Hindavi Publishing. With permission.)

(integrating comparator). The digital output signal ($V_{an\_out}$) and clock (clk) control the logic circuitry that sets bits $b_1 - b_n$, using a successive approximation algorithm.

Converter functionality is illustrated in Figure 11.11 for operation at 1-MHz sampling frequency and $I_{tr}$ current equal to 1 nA. The waveform depicts $V_c$ voltage at the input of the first NOT gate of the comparator during the SAR operation. The transistor's dimensions of the first NOT gate (comparator) were selected as minimal to minimize gate-to-source capacitance ($C_{GS}$) in order to obtain the highest possible speed of this circuit.

One of the advantages of the proposed ADC is that good matching between transistors in pairs M1-M2, M3-M4, M5-M6, M7 and M81-M88, M9-M10, M11-M12 is not critical, because any mismatch in these pairs can be compensated by adjusting the $I_{tr}$ current during device calibration. However, sufficiently good matching between transistors M81-M88 is important, as it can substantially affect the linearity of the ADC.

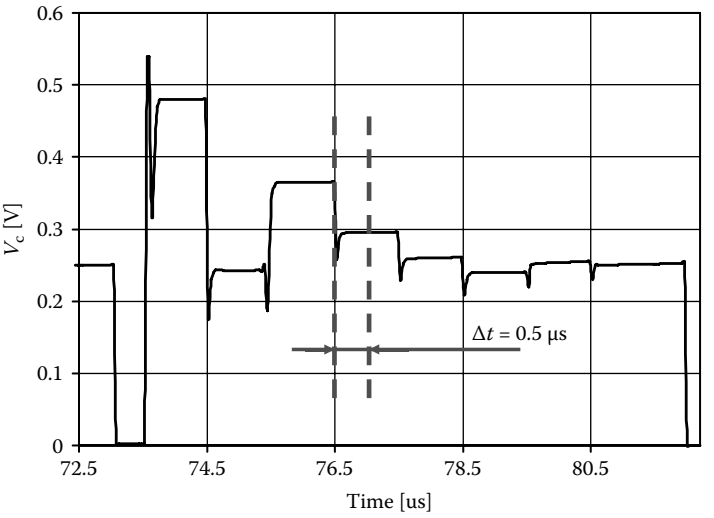



**FIGURE 11.11** Successive calculation of the output bits for an example input current $I_{in} = 211$ nA sampled with $f_s = 1$ MHz, and for $I_{tr} = 1$ nA. $V_c$ depicts voltage of the current comparator. (From Dlugosz and Iniewski [32]. © 2007 by Hindavi Publishing. With permission.)

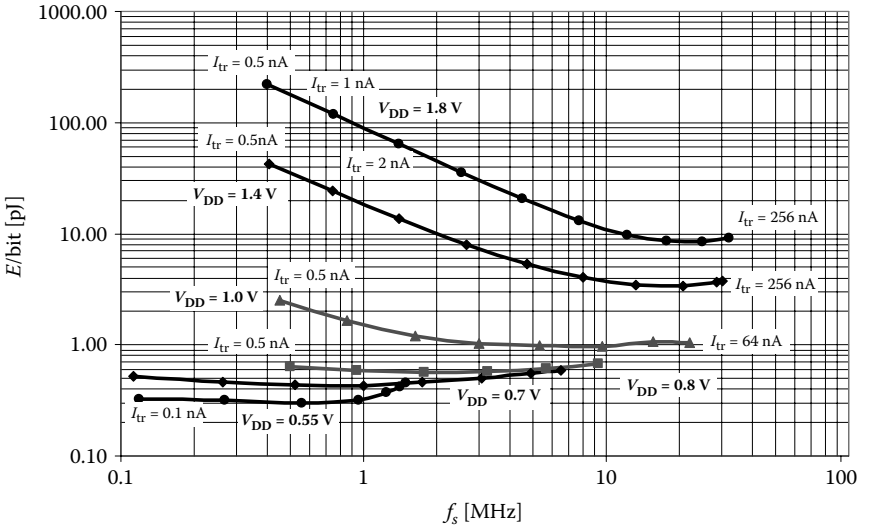



**FIGURE 11.12** Energy per calculated bit versus sampling frequency for different values of supply voltage $V_{DD}$.

The described converter has one important characteristic: Adjustment of the $I_{tr}$ current and the supply voltage $V_{DD}$ allows for the flexible regulation of power consumption and sampling frequency over a wide range. Figure 11.12 illustrates energy consumed per 1 bit versus the sampling frequency for different values of power supply ($V_{DD}$). As can be observed, the optimal region of operation depends on the $V_{DD}$ value.

The converter power is dissipated by the comparator, by the control logic block, and by the DAC. Power of the control-logic block does not exceed 10% of the overall power dissipation and increases linearly with the sampling frequency. The percentage contribution to the overall power dissipation in the case of comparators, implemented here as an inverter, depends on voltage supply. In CMOS gates working in standard digital applications, transition states between the logical values 0 and 1 are short, and energy lost during switching is relatively low. Both the input and the output voltages of the first inverter in the comparator are close to the $V_{DD}/2$ point, especially when, during the following algorithm steps, the reference signal becomes close to the input signal, as shown in Figure 11.11. When the supply voltage is low, then both transistors in the inverter operate in the weak inversion, and current that flows in the inverter can be neglected. Otherwise, when $V_{DD} > V_{TH\_pMOS} + V_{TH\_nMOS}$, then current flowing in the inverter becomes significant for the output voltage in the midpoint between $V_{DD}$ and $V_{SS}$. For small supply voltages, power dissipation linearly increases with the sampling frequency, and energy per bit is approximately constant. When supply voltages are high, then power dissipated in the comparator becomes a dominant component, especially for small input and reference currents. In this situation, power dissipated in DAC can be neglected. For small sampling frequencies and high supply voltages, power dissipation is almost constant, and energy per 1 bit increases linearly when sampling frequency decreases.

Each curve in Figure 11.12 has a minimum, i.e., the most optimal value of energy per conversion step. Figure 11.13 illustrates placement of these optimum values for particular values of the power supply vs. sampling frequency. The best parameters are attained for moderately low supply voltages around 1 V. For this supply, the optimum parameters are obtained for the sampling frequency of 10 MHz.

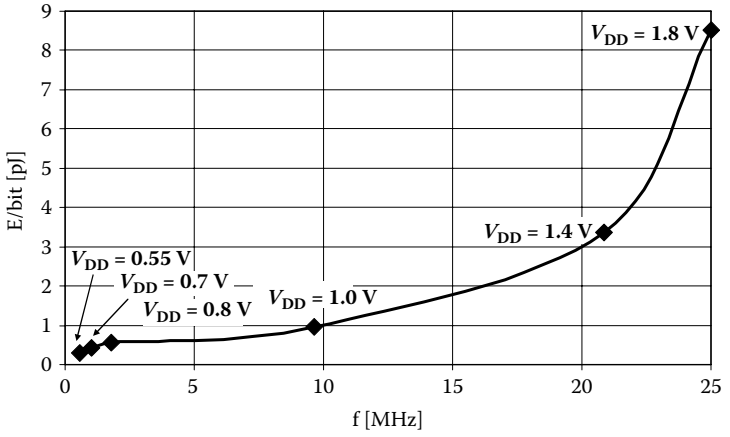



**FIGURE 11.13** Optimal energy per 1-bit versus sampling frequency for given supply voltages $V_{DD}$.

## 11.7 OPTIMIZATION TECHNIQUES IN CURRENT-MODE SAR ADCs

A problem that must be considered in the current-mode SAR ADC is transistor mismatch, which requires very precise DAC calibration, as described by Yang and Van der Spiegel [60]. The mismatch problem is especially important when transistors operate in the weak inversion region at small power supplies. In such converters, DAC is implemented as a multi-output current mirror with binary-weighted transistors, and in this case, mismatch has a direct influence on the differential nonlinearity (DNL) of the input–output characteristic of the converter. Increasing transistor size minimizes this effect, but this increases the chip area, thus limiting the maximum data resolutions in practice to only 7–8 bits.

The other technique that can be used to solve this problem has been described by Długosz, Gaudet, and Iniewski [29]. In this case, DAC has been realized as a cascade connection of two multistage current mirrors, where transistors do not need to be binary weighted. As a result, a smaller spread between transistor sizes and smaller chip area can be achieved.

In the classic single-stage binary-weighted approach, where the converter has the resolution of $n$ bits, the gain $k_x$ of the $x^{\text{th}}$ branch in DAC is equal to $2^{x-1}$. The DAC output reference current $I_{\text{ref}}$ is described as

$$I_{\text{ref}} = \sum_{x=1}^{n} I_{\text{ref}_x}, \text{ where } I_{\text{ref}_x} = k_x I_{\text{tr}} = 2^{x-1} I_{\text{tr}} \tag{11.4}$$

In the approach shown in Figure 11.14 the $I_{\text{tr}}$ current is copied $n$ times in the first DAC stage, creating $I_{\text{trx}}$ currents, which are then gained in the second stage, creating $I_{\text{refx}}$ currents. Particular reference currents are in this case given as

$$I_{\text{ref}_x} = k_{x1} k_{x2} I_{\text{tr}} \Rightarrow k_{x1} k_{x2} = 2^{x-1} \tag{11.5}$$

where $k_{x1}$ and $k_{x2}$ are gains of particular branches. A proper selection of coefficients $k_{x1}$ and $k_{x2}$ in each DAC enables significant minimization of the spread between transistor sizes that allows for increasing the size of all transistors. This technique is similar to that described by Abdelhalim, MacEachern, and Mahmoud [40] for a charge-redistribution voltage-mode SAR ADC, presented in the previous section. This approach has several advantages. It minimizes the mismatch effect, improves linearity of DAC, and allows for higher data resolutions. To obtain, for example, the resolution of 10 bits in a single-stage approach, the spread between transistor widths must be equal to 512. In the second approach, when gains of particular branches are properly selected, the maximum spread between transistor widths is only 32. Example gains of particular branches in both DAC stages can be selected as follows:

$$k_{x1} = \{1, 1, 1, 1, 1, 2, 4, 8, 16, 16\} \text{ and } k_{x2} = \{1, 2, 4, 8, 16, 16, 16, 16, 16, 32\} \tag{11.6}$$

Assuming that the width of the smallest transistor in DAC is considered to be a unity width (UW), the sum of all transistor widths for parameters given in Equation (11.6) is equal to 178·UW, in comparison to 1023·UW in the previous approach.

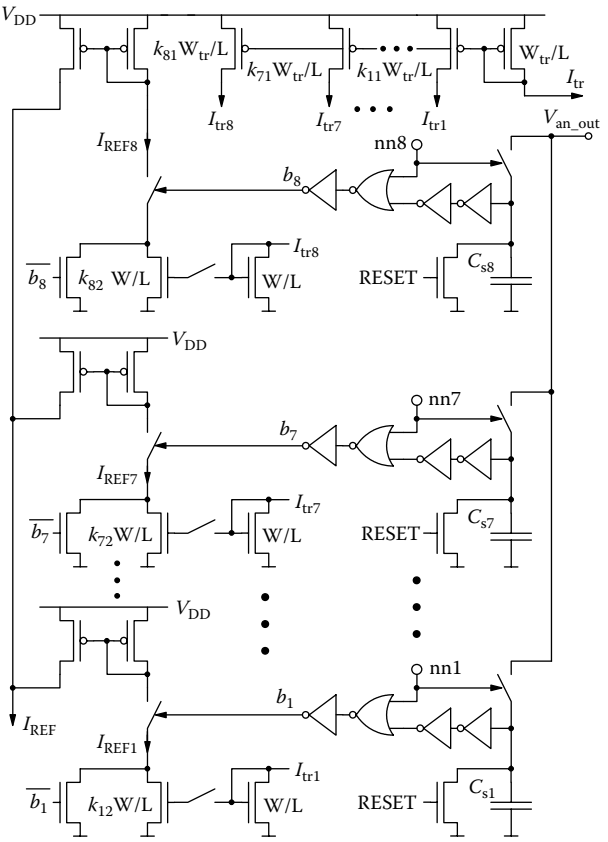



**FIGURE 11.14** Two-stage DAC with a controlling digital circuitry. (From Długosz, Gaudet, and Iniewski [29]. © 2007 by IEEE. With permission.)

This allows for increasing the value of UW and for minimizing the mismatch error. In this approach, each DAC stage introduces its own mismatch error. Assuming that the value of UW in both approaches is equal, the resultant error in the second approach will be larger by 5–10% than in the first one. On the other hand, increasing the value of UW 5.75 times (i.e., 1024/178), which is possible due to lower spread, lowers this error by up to 80%.

One of the disadvantages of the second approach is the increased number of branches in both DAC stages, resulting in slightly increased total current. The additional current is very small. For the example gains given in Equation (11.6), this is maximum $51 \cdot I_{tr}$, which is only ca. 5% of the maximal value of the reference current $I_{ref}$.

## 11.8 INTERLEAVED SAR OPERATION

A serious limitation of classic SAR ADCs is a relatively low data-conversion rate, which is equal to $f_S/n$. To overcome this problem, the interleaved SAR ADCs have

been introduced and are gaining popularity in recent years, especially in applications where a high data rate is required, e.g., in Ultra-Wideband (UWB) systems.

The general scheme of the interleaved converter is shown in Figure 11.15. This circuit consists of an analog demultiplexer at the input, a digital multiplexer at the output, and several equal channels working in parallel, where each of them contains a single classic SAR converter. A high-rate input analog data stream is directed to particular channels using the input demultiplexer, which is akin to a rotating switch, and is controlled by a central clock generator. In this way, each channel gets only a fraction of the input data and therefore may be sampled with reduced sampling frequency. Particular channels start conversion in different phases of the central clock generator and finish it also in different time moments. The results are combined into one output digital data stream using another rotating switch (multiplexer).

Particular channels operate at reduced frequencies, and therefore the interleaved approach puts lower demands on particular building blocks such as comparators and DAC, thus simplifying the design process for these blocks. On the other hand, an additional central control block is required to synchronize all channels in time, which makes the interleaved structure much more complex than the classic SAR converter. Interleaved converters occupy the chip area that is at least $n$ times bigger than in the case of a single converter.

Interleaved SAR converters, in comparison to flash converters, have several interesting features. One of them is high flexibility, meaning that their power dissipation and resolution may be easily controlled and matched to temporary conditions and system requirements. This is possible, as an SAR algorithm can be stopped at any time, and particular imaging channels may be turned off, when the input data rate is smaller.

In recent years, several interesting interleaved SAR converters have been reported, mostly designed using the voltage-mode charge-redistribution approach for UWB applications, where maximum resolution of 5 or 6 bits is sufficient, but where power dissipation is a critical criterion. These converters operate with very high sampling

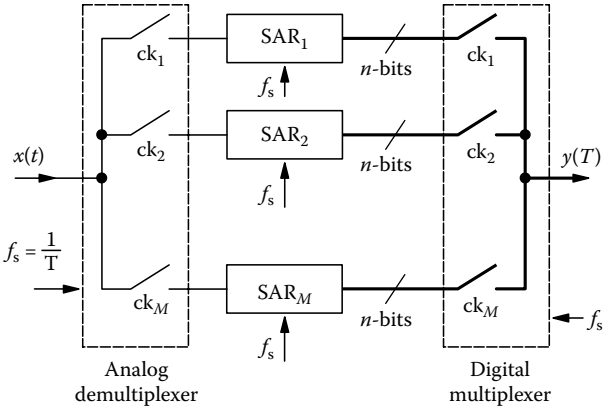



**FIGURE 11.15** The general idea of SAR interleaved ADC.

frequencies, between 500 MHz and 1.2 GHz, which so far were reserved rather for flash converters, achieving much lower power dissipation.

The relatively low resolution is mostly imposed by a capacitor array. Increasing the resolution by each additional bit means, in practice, almost doubling the chip area and power dissipation. The other reason for the low resolution is the comparator's offset. This problem can be solved by raising the input voltage swing, but this also raises the power dissipation of DAC.

Optimization techniques that usually are used in interleaved SAR converters to achieve better parameters are similar to those used in the single SAR converters we described earlier. Optimization of the capacitor array in DAC is one of these techniques. In general, the effective converter's resolution depends mostly on the linearity of the DAC, which depends on matching between capacitors. A typical method that relies on increasing the unity capacitor in DAC limits the maximum sampling frequency. This effect can be compensated by larger switches with lower on-resistance used in DAC, but this also enlarges the charge-injection effect. In the solution presented by Dondi et al. [45], this problem has been solved by using a boosting charge pump, which raises a gate-to-source voltage in switches, thus decreasing the on-resistance. This technique allows for the use of transistors with smaller widths and a reduction of the charge-injection effect. As a result, a sampling frequency up to 1.2 GHz is possible at the cost of 16 mW of power dissipation.

The other way to increase the conversion data rate is to introduce various self-timing techniques. In general, SAR converters do not require a very clean clock generator. The only task of the clock circuit is to start particular algorithm steps after the previous steps are completed. In this situation, time allocated for each step must be sufficient to settle the comparator's output. SAR converters typically are controlled by synchronous clocks, where the sampling frequency is adjusted to the worst-case scenario, i.e., when both comparator inputs have similar values and the resolving time is long. Such a situation might occur, for example, during the calculation of LSB. This approach is a source of time redundancy in other successive algorithm steps, which inherently limits the data-conversion rate. For example, one of these self-timing techniques has been used in the converter reported by Ginsburg and Chandrakasan [46]. In SAR conversion schemes, the comparative results from one step are used to prepare a new reference signal for the next step. A disadvantage of SAR architecture is that preparation of this new reference signal requires some additional feedback time. This time must be minimized to allocate as much time as possible to settle the analog signals in the capacitor array and preamplifier stages in the comparator. The self-timing technique described by Ginsburg and Chandrakasan [46] relies on minimizing this time. In this case, after the comparator resolves a value, the output latch triggers the next algorithm stage. In this situation, the remainder of the time initially allocated for the previous step may be allocated to the next step. This mechanism enables the extension of the settling time by about 20%.

The other self-timing technique has been reported by Chen and Brodersen [48], where an asynchronous clock generator was used to control a 6-bit interleaved SAR

converter. The presented technique allocates only as much time to each algorithm step as required to settle the comparator's output. When both comparator inputs differ significantly, then the comparator settles very quickly, and the next algorithm step can start immediately. This concept requires a special dynamic comparator that can generate a data-ready signal when comparison is completed. The comparator used in this converter has two complementary outputs, which during the reset phase are connected to the positive supply voltage, generating logical values 11. During resolving time, one of these outputs becomes 0. The data-ready signal is generated by an additional logic circuit, which distinguishes a 01 or 10 state from the previous 11 state. This technique is very effective, resulting in the best parameters between the converters described in this section. The converter dissipates 5.3 mW at a sampling frequency of 600 MHz.

A new interleaved 6-bit ADC suitable for high sampling frequencies above 1 GS/s and moderate power dissipation has recently been proposed by Cao, Yan, and Li [57]. In this converter, shown in Figure 11.16, two equal structures are interleaved in time and clocked with the sampling frequency of 1.25 GS/s each. This is a hybrid structure that involves both the flash architecture and the SAR algorithm. The top conversion scheme is the SAR one, while at each conversion stage a 2-bit flash converter is used. As a result, to calculate 6 bits, only 3 conversion steps are required. This converter is a compromise solution between the SAR and the flash structures. Instead of using 63 comparators, as would be the case in a 6-bit flash converter, in this case, the number of converters has been reduced to only 6, i.e., 3 for each channel. The power dissipation at the level of 32 mW is in this case higher than in previously described SAR converters, but it is still five to six times smaller than in

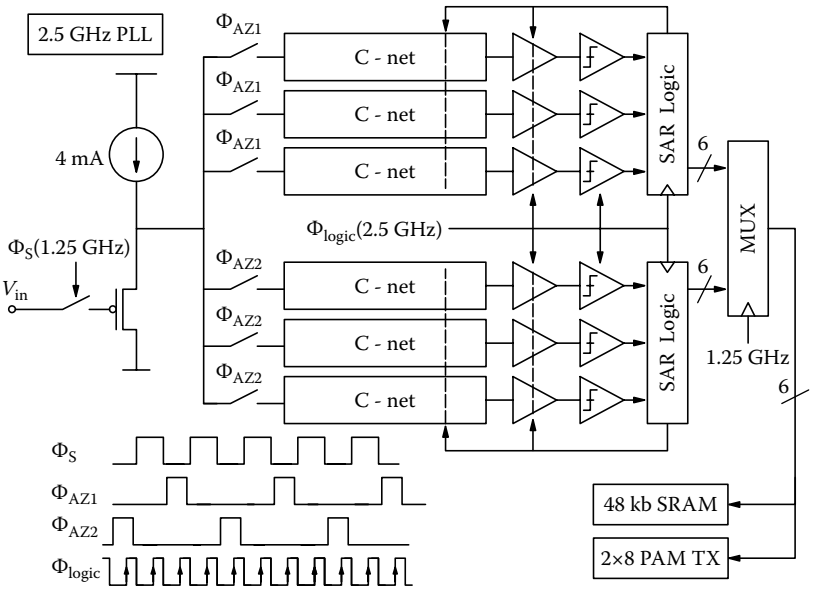



**FIGURE 11.16** Hybrid flash-SAR high-speed ADC. (From Cao, Yan, and Li [57]. © 2009 by IEEE. With permission.)

classic flash structures operating with the same frequencies [9]. The disadvantage here is the increased number of capacitors. However, a careful design of the capacitor array (C-net) allowed reducing the value of the UC to only 5 fF, which condensed the overall chip area to a relatively small 0.09 $mm^2$.

## 11.9 CURRENT-MODE INTERLEAVED SAR ADC

All interleaved converters described in the previous section have been designed for very high sampling frequencies in order to compete with flash architectures. Comparing these converters with flash counterparts designed for the same sampling frequency range [6, 8–10], we can see that the power dissipation in the interleaved SAR converters is about one order of magnitude smaller, while the chip area in both approaches is comparable. In most medical-imaging applications, a higher circuit complexity that offers lower power dissipation is fully justified.

The interesting question is whether there is any sense in using the interleaved SAR approach in the case of sampling frequencies, which can be easily attained by single SAR converters that feature a simpler structure and lower chip area. The answer in some cases is positive. When looking, for example, at the current-mode converter described in Section 11.4 and at the results shown in Figure 11.13, the conclusion is as follows. When higher sampling frequencies are required (e.g., 80 MHz) and when the chip area is not a critical parameter, then an interleaved SAR structure with eight channels working in parallel is a more efficient solution. In this case, the interleaved converter operating at 1-V supply consumes 10 pJ/bit at the sampling frequency of 80 MHz, while a single SAR converter operating at 1.8 V consumes a similar energy of about 9 pJ/bit, but for a smaller sampling frequency of 25 MHz.

An example current-mode interleaved SAR converter has been reported by Dlugosz and Iniewski [32]. This converter consists of eight parallel channels, each containing a single SAR block described in Section 11.4. This converter has been designed for applications where the input data rate can vary over a wide range, and where power dissipation is a main criterion. The central control-logic block used in this circuit puts those sections that are temporarily not used into the power-down mode. In this case, the circuit consumes only a small fraction of the energy it would otherwise use. In this ADC converter, particular sections wake up only when they receive the analog data for conversion, and are turned off immediately after completing the conversion task. This is illustrated in Figures 11.17 and 11.18. Power dissipation starts increasing when the first ADC section wakes up. Maximum current is reached when all eight sections are turned on and then decreases when the subsequent sections complete their conversion tasks. This converter implemented in TSMC 0.18-μm CMOS technology occupies a 0.1-$mm^2$ area and, at the sampling frequency of 1 MHz, dissipates 4 μW of power from a 0.65-V power supply. In standby mode, when all channels are off, the average power dissipation is 130 nW, consumed mostly by a central logic block.

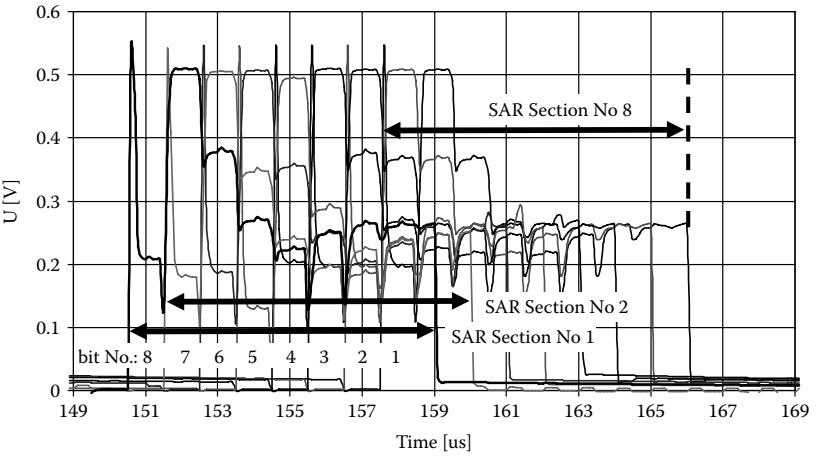



**FIGURE 11.17** Voltage waveforms at the comparator inputs, illustrating the interleaving action of eight SAR sections sequentially turning on and off. (From Dlugosz and Iniewski [32]. © 2007 by Hindavi Publishing. With permission.)

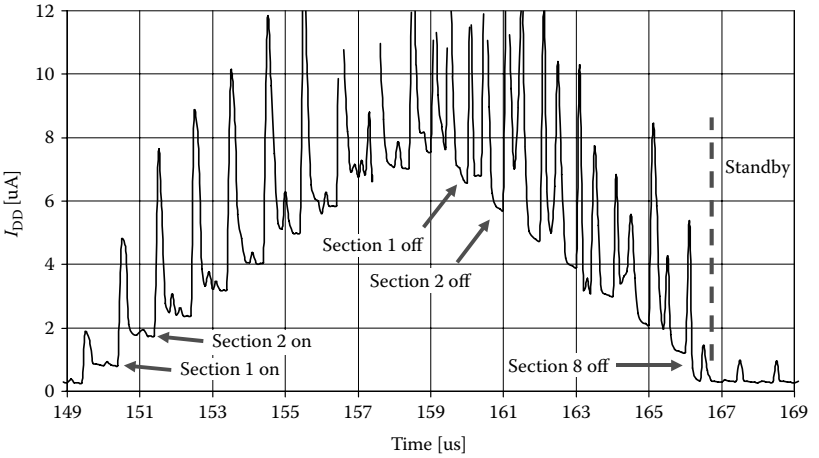



**FIGURE 11.18** Total current flowing in the converter during particular conversion steps and during the standby mode. (From Dlugosz and Iniewski [32]. © 2007 by Hindavi Publishing. With permission.)

## 11.10 CONCLUSIONS

X-ray medical-imaging sensors integrated with traditional CMOS circuitry are poised to enable a new era of digital imaging where images gathered using various imaging techniques will be processed, stored, and transmitted using well-known digital media. To achieve cost effectiveness of these techniques, a significant research effort is required both in imaging sensors and in the accompanying CMOS circuits. One

of the critical areas that requires further innovation is analog-to-digital conversion. In this chapter, we have reviewed different ADC design architectures and suggested that successive approximation (SAR) is the most power-efficient design technique.

A larger selection of SAR-based stand-alone commercial products is expected in the future, which will enable more power-efficient and integrated imaging systems. On-chip ADC integration with analog front-end circuitry that interfaces directly to X-ray sensors is also possible. However, common problems with noise cross talk, circuit thermal and flicker noise, manufacturing variations, and mismatch and low-voltage signal processing will have to be addressed in CMOS-chip implementations.

## REFERENCES


1. van de Plassche, Rudy J. 2003. *CMOS integrated analog-to-digital and digital-to-analog converters*. Boston: Kluwer.
2. Abrial, A., J. Bouvier, J.-M. Fournier, and P. Senn. 1993. A low-power 8-b 13.5-MHz video CMOS ADC for visiophony ISDN applications. *IEEE Journal of Solid-State Circuits* 28:725–729.
3. Hosotani, S., T. Miki, A. Maeda, and N. Yazawa. 1990. An 8-bit 20-MS/s CMOS A/D converter with 50-mW power consumption. *IEEE Journal of Solid-State Circuits*. 25:167–172.
4. Lin, J., and B. Haroun. 2002. An embedded 0.8 V/480 µW 6 b/22 MHz flash ADC in 0.13-µm digital CMOS process using a nonlinear double interpolation technique. *IEEE Journal of Solid-State Circuits*. 37:1610–1617.
5. Terada, J., Y. Matsuya, F. Morisawa, and Y. Kado. 2000. 8-mW, 1-V, 100-MSps, 6-bit A/D converter using a transconductance latched comparator ASICs. In *Proceedings of the Second IEEE Asia Pacific Conference on ASICs* (AP-ASIC), 53–56.
6. Sutardja, P., D. D. Tang, J. Altieri, L. E. Thon, G. Coleman, S. Subbanna, and J. Y.-C. Sun. 1995. 55-mW 300-MHz analog-digital converters using digital VLSI technology. In *Proceedings of IEEE Symposium on Low Power Electronics*, 68–69.
7. Culurciello, E., and A. Andreou. 2003. An 8-bit, 1 mW successive approximation ADC in SOI CMOS. In *Proceedings of the International Symposium on Circuits and Systems* (ISCAS) 1:I-301–I-304.
8. Hadji-Abdolhamid, A., and D. A. Johns. 2003. A 400-MHz 6-bit ADC with a partial analog equalizer for coaxial cable channels. In *Proceedings of the 29th European Solid-State Circuits Conference* (ESSCIRC), 237–240.
9. Sandner, C., M. Clara, A. Santner, T. Hartig, and F. Kuttner. 2005. A 6-bit 1.2-GS/s low-power flash-ADC in 0.13-µm digital CMOS. In *Proceedings of IEEE Journal of Solid-State Circuits* 40:1499–1505.
10. Geelen, G., and E. Paulus. 2004. An 8 b 600 MS/s 200 mW CMOS folding A/D converter using an amplifier preset technique. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC) 1:254–256.
11. Draxelmayr, D. 2004. A 6b 600MHz 10mW ADC array in digital 90 nm CMOS. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC) 1:264–527.
12. Bright, W. 1998. 8 b 75 M sample/s 70 mW parallel pipelined ADC incorporating double sampling. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 146–147, 428.

13. Kim, Se-Won, Young-Jae Cho, Kyung-Hoon Lee, Seung-Hoon Lee, Jae-Yup Lee, Hyun-Chul Noh, and Hee-Sub Lee. 2005. An 8 b 240 MS/s 1.36 mm$^2$ 104 mW 0.18 μm CMOS ADC for DVDs with dual-mode inputs. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS), 4054–4057.
14. Koo, Ja-Hyun, Yun-Jeong Kim, Sin-Hu Kim, Won-Joo Yun, Shin-Il Lim, and Suki Kim. 2004. An 8-bit 250MSPS CMOS pipelined ADC using open-loop architecture. In *Proceedings of the IEEE Asia-Pacific Conference on Advanced System Integrated Circuits* (AP-ASIC), 94–97.
15. Mulder, J., C. M. Ward, Chi-Hung Lin, D. Kruse, J. R. Westra, M. L. Lugthart, E. Arslan, R. van de Plassche, K. Bult, and F. M. L. van der Goes. 2004. A 21mW 8b 125MS/s ADC occupying 0.09 mm$^2$ in 0.13 μm CMOS. Paper presented at IEEE International Solid-State Circuits Conference (ISSCC), 15–19 February 2004.
16. Nishida, Y., K. Sone, K. Amano, S. Matsuba, and A. Yukawa. 1995. An 8-bit 200 MS/s 500 mW BiCMOS ADC. In *Proceedings of the IEEE Custom Integrated Circuits Conference* (CICC), 207–210.
17. Sigenobu, T., M. Ito, and T. Miki. 2001. An 8-bit 30 MS/s 18 mW ADC with 1.8 V single power supply. In *Proceedings of Symposium on VLSI Circuits*, 209–210.
18. Kim, Kyung Myun, and Kwang Sub Yoon. 1996. An 8-bit 42 Msamples/s current-mode folding and interpolation CMOS analog-to-digital converter with three-level folding amplifiers. In *Proceedings of IEEE 39th Midwest Symposium on Circuits and Systems* 1:201–204.
19. Nauta, B., and A. G. W. Venes. 1995. A 70 Msample/s 110 mW 8 b CMOS folding interpolating A/D converter. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 276–277, 379.
20. Agarwal, A., Y. B. Kim, and S. Sonkusale. 2005. Low power current mode ADC for CMOS sensor IC. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS) 1:584–587.
21. Taft, R. C., C. A. Menkus, M. R. Tursi, O. Hidri, and V. Pons. 2004. A 1.8-V 1.6-Gsample/s 8-b self-calibrating folding ADC with 7.26 ENOB at Nyquist frequency. *IEEE Journal of Solid-State Circuits* 39:2107–2115.
22. Lee, Jungeun, and Minkyu Song. 1999. Design of an 8-bit 100 kSps 1 mW CMOS A/D converter for digital mobile communication. In *Proceedings of the First IEEE Asia Pacific Conference on ASICs* (AP-ASIC), 178–185.
23. Kim, Seong-Won, and Soo-Won Kim. 1991. Current-mode cyclic ADC for low power and high speed applications. *Electronics Letters* 27:818–820.
24. Tanner, S., A. Heubi, M. Ansorge, and F. Pellandini. 1998. An 8-bit low-power ADC array for CMOS image sensors. *IEEE International Conference on Electronics Circuits and Systems* (ICECS) 1:147–150.
25. Cauwenberghs, G. 1995. A micropower CMOS algorithmic A/D/A converter. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* [see also *IEEE Transactions on Circuits and Systems I: Regular Papers*] 42:913–919.
26. Scott, M. D., B. E. Boser, and K. S. J. Pister. 2003. An ultralow-energy ADC for Smart Dust. *IEEE Journal of Solid-State Circuits* 38:1123–1129.
27. Ravezzi, L., D. Stoppa, and G.-F. Dalla Betta. 1998. Current-mode A/D converter. *Electronics Letters* 34:615–616.
28. Razavi, B. 1995. *Principles of data conversion system design*. New York: IEEE Press.
29. Długosz, R., V. Gaudet, and K. Iniewski. 2007. Flexible ultra low power successive approximation analog-to-digital converter with asynchronous clock generator. In *Proceedings of Canadian Conference on Electrical and Computer Engineering* (CCECE), 1649–1652.

30. Scott, M. D., B. E. Boser, and K. S. J. Pister. 2002. An ultra-low power ADC for distributed sensor networks. In *Proceeding of the 28th European Solid-State Circuits Conference* (ESSCIRC), 255–258.
31. Verma, N., and P. Chandrakasan. 2006. A 25 μW 100 kS/s 12b ADC for wireless micro-sensor applications. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 222–223.
32. Dlugosz, R., and K. Iniewski. 2007. Ultra low power current-mode algorithmic analog-to-digital converter implemented in 0.18 μm CMOS technology for wireless sensor network. In *VLSI Design*. New York: Hindavi Publishing. DOI:10.1155/2007/45269.
33. Lin, Chi-Sheng, and Bin-Da Liu. 2003. A new successive approximation architecture for low-power low-cost CMOS A/D converter. *IEEE Journal of Solid-State Circuits* 38:54–62.
34. Sauerbrey, J., D. Schmitt-Landsiedel, and R. Thewes. 2003. A 0.5-V 1 μW successive approximation ADC. *IEEE Journal of Solid-State Circuits* 38:1261–1265.
35. Yoshida, T., M. Akagi, M. Sasaki, and A. Iwata. 2005. A 1V supply successive approximation ADC with rail-to-rail input voltage range. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS). 1:192–195.
36. Borghetti, F., J. H. Nielsen, V. Ferragina, P. Malcovati, P. Andreani, and A. Baschirotto. 2006. A programmable 10 b up-to-6 MS/s SAR-ADC featuring constant-FoM with on-chip reference voltage buffers. In *Proceedings of the 32nd European Solid-State Circuits Conference* (ESSCIRC), 500–503.
37. Verma, N., and A. P. Chandrakasan. 2007. An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor nodes. *IEEE Journal of Solid-State Circuits* 42:1196–1205.
38. Craninckx, J., and G. van der Plas. 2007. A 65 fJ/conversion-step 0-to-50 MS/s 0-to-0.7 mW 9 b charge-sharing SAR ADC in 90 nm digital CMOS. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 246–247.
39. Aksin, D., M. A. Al-Shyoukh, and F. Maloberti. 2007. An 11 bit sub-ranging SAR ADC with input signal range of twice supply voltage. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS), 1717–1720.
40. Abdelhalim, K., L. MacEachern, and S. Mahmoud. 2007. A nanowatt successive approximation ADC with offset correction for implantable sensor applications. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS), 2351–2354.
41. Arbat, A., A. Dieguez, and J. Samitier. 2006. An ultra low power successive approximation ADC with selectable resolution in 0.13 um CMOS technology. In *Proceedings of 13th IEEE International Conference on Electronics, Circuits and Systems* (ICECS), 415–418.
42. Robert, Pierre-Yves, Benoit Gosselin, Amer Elias Ayoub, and Mohamad Sawan. 2006. An ultra-low-power successive-approximation-based ADC for implantable sensing devices. In *Proceedings of 49th IEEE International Midwest Symposium on Circuits and Systems* (MWSCAS) 1:7–11.
43. Hesener, M., T. Eicher, A. Hanneberg, D. Herbison, F. Kuttner, and H. Wenske. 2007. A 14 b 40 MS/s redundant SAR ADC with 480 MHz clock in 0.13 μm CMOS. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 248–600.
44. Ng, Sheiing Yan, Bahar Jalali, Pengbei Zhang, J. Wilson, and Mohammad Ismail. 2005. A low-voltage CMOS 5-bit 600 MHz 30 mW SAR ADC for UWB wireless receivers. In *Proceedings of 48th Midwest Symposium on Circuits and Systems* (MWSCAS) 1:187–190.
45. Dondi, S., D. Vecchi, A. Boni, and M. Bigi. 2006. A 6-bit, 1.2 GHz interleaved SAR ADC in 90nm CMOS. In *Proceedings of Research in Microelectronics and Electronics, 2006, Ph.D.*, 301–304.

46. Ginsburg, B. P., and A. P. Chandrakasan. 2007. Dual time-interleaved successive approximation register ADCs for an ultra-wideband receiver. *IEEE Journal of Solid-State Circuits* 42:247–257.
47. Ginsburg, B. P., and A. P. Chandrakasan. 2007. 500-MS/s 5-bit ADC in 65-nm CMOS with split capacitor array DAC. *IEEE Journal of Solid-State Circuits* 42:739–747.
48. Chen, S.-W. M., and R. W. Brodersen. 2006. A 6-bit 600-MS/s 5.3-mW asynchronous ADC in 0.13 μm CMOS. *IEEE Journal of Solid-State Circuits* 41:2669–2680.
49. Ahmed, I., and D. A. Johns. 2005. A 50-MS/s (35 mW) to 1-kS/s (15 μW) power scaleable 10-bit pipelined ADC using rapid power-on opamps and minimal bias current variation. *IEEE Journal of Solid-State Circuits* 40:2446–2455.
50. Chang, You-Kuang, Chao-Shiun Wang, and Chorng-Kuang Wang. 2007. A 8-bit 500-KS/s low power SAR ADC for bio-medical applications. In *Proceedings of IEEE Asian Solid-State Circuits Conference* (ASSCC), 228–231.
51. Roy, Tan Kuo Hwi, and T. H. Teo. 2007. 0.9 V 100 nW rail-to-rail SAR ADC for bio-medical applications. In *Proceedings of International Symposium on Integrated Circuits* (ISIC), 481–484.
52. Agnes, A., E. Bonizzoni, P. Malcovati, and F. Maloberti. 2008. A 9.4-ENOB 1 V 3.8 μW 100 kS/s SAR ADC with time-domain comparator. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 246–610.
53. Erario, F., A. Agnes, E. Bonizzoni, and F. Maloberti. 2008. Design of an ultra-low power time interleaved SAR converter. In *Proceedings of Ph.D. Research in Microelectronics and Electronics* (PRIME), 245–248.
54. Verma, N., and A. P. Chandrakasan. 2007. An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor nodes. *IEEE Journal of Solid-State Circuits* 42:1196–1205.
55. Gambini, S., and J. Rabaey. 2007. Low-power successive approximation converter with 0.5 V supply in 90 nm CMOS. *IEEE Journal of Solid-State Circuits* 42:2348–2356.
56. Hesener, M., T. Eichler, A. Hanneberg, D. Herbison, F. Kuttner, and H. Wenske. 2007. A 14 b 40 MS/s redundant SAR ADC with 480 MHz clock in 0.13 μm CMOS. In *Proceedings of IEEE International Solid-State Circuits Conference* (ISSCC), 248–249.
57. Cao, Zhiheng, Shouli Yan, and Yunchu Li. 2009. A 32 mW 1.25 GS/s 6 b 2 b/step SAR ADC in 0.13 μm CMOS. *IEEE Journal of Solid-State Circuits* 44:862–873.
58. Singh, Pratap Narayan, Ashish Kumar, Chandrajit Debnath, and Rakesh Malik. 2008. A 1.2V 11 b 100 MSps 15 mW ADC realized using 2.5 b pipelined stage followed by time interleaved SAR in 65 nm digital CMOS process. In *Proceedings of IEEE Custom Integrated Circuits Conference* (CICC), 305–308.
59. Hong, Hao-Chiao, and Guo-Ming Lee. 2007. A 65-fJ/conversion-step 0.9-V 200-kS/s rail-to-rail 8-bit successive approximation ADC. *IEEE Journal of Solid-State Circuits* 42:2161–2168.
60. Yang, Zheng, and J. Van der Spiegel. 2008. A 10-bit 8.3 MS/s switched-current successive approximation ADC for column-parallel imagers. In *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS), 224–227.
61. Ali A.M.A., Dillon C., Sneed R., Morgan A.S., Bardsley S., Kornblum J., Lu Wu. 2006. "A 14-bit 125 MS/s IF/RF Sampling Pipelined ADC With 100 dB SFDR and 50 fs Jitter," *IEEE Journal of Solid-State Circuits*, 41:1846–1855.
62. Deguchi, K., N. Suwa, M. Ito, T. Kumamoto, and T. Miki. 2008. A 6-bit 3.5-GS/s 0.9-V 98-mW flash ADC in 90-nm CMOS. *IEEE Journal of Solid-State Circuits* 43:2303–2310.
63. Park, Sunghyun, Y. Palaskas, and M. P. Flynn. 2007. A 4-GS/s 4-bit flash ADC in 0.18-μm CMOS. *IEEE Journal of Solid-State Circuits* 42:1865–1872.

64. Lin, Ying-Zu, Yen-Ting Liu, and Soon-Jyh Chang. 2007. A 5-bit 4.2-GS/s flash ADC in 0.13-μm CMOS. In *Proceedings of IEEE Custom Integrated Circuits Conference* (CICC), 213–216.
65. Kertis, R.A., et al. 2008. A 35 GS/s 5-bit SiGe BiCMOS flash ADC with offset corrected exclusive-or comparator. In *Proceedings of IEEE Bipolar/BiCMOS Circuits and Technology Meeting* (BCTM), 252–255.
66. Meguellati, S., and N. E. Bouguechal, R. Arnold, and O. Manck. 2005. A charge redistribution SAR ADC for a pressure correction ASIC. In *Proceedings of 12th IEEE International Conference on Electronics, Circuits and Systems* (ICECS), 1–4.
67. Yousif, A. S., and J. W. Haslett. 2007. A fine resolution TDC architecture for next generation PET imaging. *IEEE Transactions on Nuclear Science* 54:1574–1582.
68. Swann, B. K., et al. 2004. A 100-ps time-resolution CMOS time-to-digital converter for positron emission tomography imaging applications. *IEEE Journal of Solid-State Circuits* 39:1839–1852.
69. Dlugosz R., Talaska T., "Low Power Current-Mode Binary-Tree Asynchronous Min/Max Circuit," *Microelectronics Journal*, Elsevier, Vol. 41, No. 1, January 2010.

# 12 Low-Power Integrated Front-End for Timing Applications with Semiconductor Radiation Detectors

*Sorin Martoiu and Angelo Rivetti*

## CONTENTS



## 12.1 INTRODUCTION

In several applications it is important to perform an accurate measurement of the time at which a given hit occurs. A common example is provided by time-of-flight (TOF) techniques employed for particle identification in high-energy physics experiments (Green 2000). Time-of-flight telescopes based on semiconductor radiation

detectors have been used for decades in nuclear physics (Spieler 1982). In these systems, a thin device senses the passage of a particle and generates a start signal, while a thick sensor stops the particle and measures its energy. An accurate measurement of the time elapsed between the two events combined with the energy information leads to the particle mass. The number of channels now in use for similar applications in low- and medium-energy nuclear physics justifies the need of high-density integrated front-end electronics incorporating on chip both energy and timing capabilities. A 16-channel chip allowing an energy resolution of 38 keV and a time resolution of 1-ns FWHM (full width at half maximum) has recently been reported (Engel et al. 2007).

In the field of medical imaging, the advances recently made in scintillators, photodetectors, and electronics have renewed the interest in the use of time-of-flight information in positron emission tomography (PET) (Moses 2007). Here the accurate knowledge of the time at which the photons created by the annihilation of a positron–electron pair impact the detector allows for a significant reduction of the statistical noise in the reconstructed image. However, to bring an advantage that justifies the increase in complexity, the time resolution of a TOF-PET machine must be below 500-ps FWHM without sacrificing energy or space accuracy. Silicon photomultipliers that combine high gain, fast rise time, and high quantum efficiency are good candidates to play an important role in this context in the near future.

Hybrid pixel detectors with time resolution below 200 ps rms are under development for the NA62 experiment at CERN. Designed to study the very rare $K^+ \rightarrow \pi^+ \nu\bar{\nu}$ decay channel at the CERN SPS accelerator, this experiment will allow some stringent tests of the standard model of particle physics, complementing measures that will be done at the Large Hadron Collider (LHC) (Fiorini et al. 2007). In this application, three stations of pixel sensors, each covering an area of 12 $cm^2$, will provide particle tracking and time stamping with a resolution adequate to separate tracks impinging on the detector within an interval of 1 ns.

These three examples taken from completely different fields illustrate the importance of high-resolution timing systems based on semiconductor radiation sensors. With the general tendency toward detectors with an increasing number of channels, it becomes more and more critical to have compact and low-power building blocks that allow the effective implementation of multichannel integrated front-end circuits with accurate timing capability. A time-measuring system consists of two fundamental parts: a "trigger" that, at the passage of the impinging radiation, provides a prompt pulse as accurate as possible and a time-to-digital converter (TDC) that translates the time difference between this pulse and a known reference into a corresponding digital word. In this chapter, we concentrate on the triggering function, focusing on its implementation in CMOS (complementary metal-oxide semiconductor) technologies in the form of compact and low-power circuits, suitable for being integrated on board of complex mixed-mode integrated circuits (ICs). Section 12.2 discusses the key factors affecting the resolution of a time-resolving system. Section 12.3 describes architectures based on constant-fraction discriminators (CFDs), while circuits performing amplitude correction are treated in Section 12.4.

## 12.2 TIME RESOLUTION

### 12.2.1 Timing Errors due to Noise

Timing measurement focuses on the determination of the time when a particular event has happened (i.e., a charged particle has crossed the detector), with a certain accuracy. The information on the energy associated with the event has a lesser priority or may even be disregarded. This fact fundamentally changes the way timing radiation systems are designed and optimized. The various factors affecting the resolution of a time-measuring system have been extensively analyzed in the literature (Knoll 2000; Spieler 1982, 2006). We provide here a concise description of the most relevant issues to facilitate the reading of the remaining parts of this chapter.

In principle, a system designed for timing measurements transforms the detector signal that corresponds to the desired event into an unambiguous digital signal, separating it from other types of events and from the background noise. The digital signal is compared to a temporal reference in a time-to-digital circuit that generates a time stamp or can be passed to a coincidence circuit that looks for the occurrence of a broader event.

In a simple example, the signal coming directly from the detector, or opportunely amplified and shaped, is continuously compared to a fixed threshold level. When the signal crosses this threshold level, the digital timing signal is issued. Such a circuit is generally known as a *leading-edge discriminator.* In the presence of noise, the threshold level should be set significantly higher than the noise level to limit the occurrence of fake signals triggered by noise. Apart from this, the presence of noise has a significant influence on the timing accuracy of the system, which can be understood with the help of Figure 12.1. As the signal level approaches the threshold, the noise fluctuation that modulates the signal may cause it to cross the threshold level sooner or later than in the ideal noiseless case. This uncertainty window increases with the noise level, but shrinks if the signal is faster. The effect is called *jitter* and is quantitatively given by the ratio between the noise and the slope of the signal.

$$\sigma_t = \frac{\sigma_n}{dV/dt} \tag{12.1}$$

This expression implies that a system for time measurement has to be optimized for maximum *slope-to-noise* ratio, as opposed to the *signal-to-noise* ratio optimization that is required for an amplitude-measuring system. This fact has important implications on the design criteria for timing systems.

Reducing noise alone does not necessarily provide a better timing resolution. Suppose that one decides to increase the integration time of the preamplifier (i.e., the amount of time the preamplifier integrates the charge coming from the detector in order to retrieve the intensity of the radiation event). This is equivalent to reducing the bandwidth of the preamplifier, averaging out more thermal noise, thus reducing the noise level. On the other hand, the signal at the output of the preamplifier is slower, so the denominator of Equation (12.1) is smaller, which, in turn, causes a degradation of the timing accuracy.

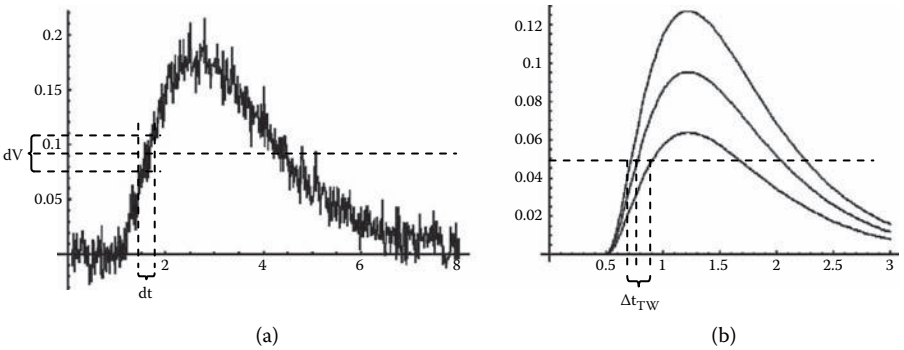



**FIGURE 12.1** Timing errors: (a) jitter; (b) time-walk.

In the other direction, if the front-end bandwidth is increased too much and the shaping time becomes faster than the detector signal, the slope of the signal does not improve significantly, as it is dominated by the charge collection time in the sensor. The noise level then rises because the higher bandwidth of the preamplifier determines an increase of its output noise (the numerator in Equation [12.1]), while the slope stays approximately the same, thereby degrading the jitter figure.

It can be shown (Spieler 2006) that, for an optimum pulse shaping the integration time in the front-end electronics should match the collection time of the detector. For a typical semiconductor sensor with thickness of 300 μm, the latter is on the order of several nanoseconds. With such short shaping times, the white noise contribution becomes dominant. This fact links the performance of the detection system to the power consumed in the preamplifier and, particularly, in the input transistor, which, in a well-designed circuit, should be the dominant noise contributor.

Another parameter that could, in principle, be manipulated to improve the timing performance is the threshold level. In fact, one can set the threshold at the exact point where the signal slope is maximum. This choice does not affect in any way the noise level, thus optimizing the time resolution. In practice, however, this is rarely possible, since the amplitude of the signal is generally variable. Furthermore, the threshold level may be imposed by the need to select a specific type of event, which prevents its use for timing optimization.

### 12.2.2 Timing Errors due to Amplitude Variations

For a leading-edge discriminator system, the amplitude variation of the signal in front of the discriminator produces another type of timing uncertainty. A front-end amplifier is usually designed to provide nomothetic output waveforms, i.e., signals corresponding to different input charges have different amplitudes, but the same shape. As shown in Figure 12.1b, since the peaking time is constant, the signal slopes change, so waveforms with higher amplitude cross the threshold earlier than the ones with lower amplitude. In a first-order approximation, the rising edge of the front-end response can be considered as a linear ramp. In this case, the slope of a waveform with amplitude $V$

is simply given by $\frac{V}{t_p}$, where $t_p$ is the peaking time. Two signals with amplitude $V_1$ and $V_2$ cross, hence the threshold $V_{th}$ at

$$t_1 = V_{th}\frac{t_p}{V_1} \text{ and } t_2 = V_{th}\frac{t_p}{V_2},$$

respectively. As it is seen from these relationships, the time at which the threshold is reached is inversely proportional to the signal amplitude. Called *time-walk*, the difference between the two crossing points represents a systematic effect of the amplitude variation of the signal. It must be said that in real situations the rising edge of the pulse shape is not a ramp, so the relationship between amplitude and crossing time is actually more complex. The consequence of this fact will be discussed in a practical example later in this chapter. Let us now examine the most common causes of amplitude variation in a detector readout system.

Charged particles crossing detector volume generate a certain number of electron–hole pairs that, traveling in the applied electric field, induce a signal on the collecting electrodes. The average amount of generated charge is obtained from the integral of the Bethe-Block formula (Leo 1987), which gives the average rate of the ionization energy loss of the particle in the detector volume. For low-momentum particles moving at nonrelativistic velocities, the energy loss rate is inversely proportional to the energy itself. Consequently, the charge signal follows the same relation. At higher energies, the energy-loss rate shows a minimum before it rises again due to relativistic effects and then saturates, owing to the polarization of the medium (Green 2000).

In high-energy physics experiments, the charged particles that cross a semiconductor detection system are mostly *minimum-ionizing particles* (MIPs). This means that the average charge signal generated by them does not change significantly with the particle's energy or species. However, even if the average deposited charge stays constant, there are statistical fluctuations in the energy-loss process, which are described by the Landau distribution (Landau 1944; Bichsel 1988). Consider as an example a monoenergetic beam of high-momentum particles impinging on a silicon detector 300-μm thick. Even if all the particles have the same kinetic energy, the deposited charge can change by a factor of three to five from one ionizing event to the other (Spieler 2006), thus determining a significant time-walk in the front-end circuit.

It must be said here that the statistical nature of the signal generation helps reduce the impact of the time-walk on the final measurement. In fact, due to shape of the Landau curve, the amplitude distribution is not uniform, but it is concentrated into a smaller amplitude range. This leads to a lower standard deviation of the time response. In the case of MIPs, for example, one can treat the time-walk as a statistical variation and use the Landau distribution to compute the resulting standard deviation. It should also be noted that semiconductor detectors are usually segmented in one (strip) or two (pixel) dimensions, so a single sensor has many independent sensing elements. When the ionizing particle crosses the sensor in between two pixels or strips, the charge generated is shared by the two neighbors, further decreasing the

available signal amplitude and increasing the dynamic range that must be considered in the system design.

The variation of the amplitude also influences the way the jitter affects the time resolution. The jitter has an inversely proportional relation to the amplitude, just as the time-walk does. This is easily understood from Equation (12.1), taking into account that the rising time of the signal remains constant, so its slope is reduced. In other words, low-amplitude signals generate considerably more uncertainty than high-amplitude signals do. As a result, the effect of jitter has to be tailored by the amplitude distribution. For example, in the case of a minimum-ionizing particle, the jitter that corresponds to the most probable amplitude in the Landau distribution can be considered, in first approximation, as the dominant component.

In applications that require a very good time resolution (i.e., significantly lower than the shaping time or the collection time of the detector), the time-walk error needs to be compensated for. In principle, the compensation methods can be classified in two categories: *off-line* methods, when enough information corresponding to the measured event is recorded and later used to reconstruct the time information with better accuracy; and *online* methods, when an auto-compensation scheme is implemented in the front-end circuitry, which then provides an accurate enough time stamp.

Given its systematic nature, the time-walk can be compensated *off-line* by recording in parallel the amplitude of the signal and correcting the time measurement in the data-analysis process. In principle, this method can fully eliminate the time-walk error, provided that several conditions are met. As a disadvantage, the amplitude measurement increases the amount of information that has to be recorded by the data acquisition system, increasing the data bandwidth as well. The accuracy of the additional amplitude measurement enters, of course, in the final time resolution along with the jitter.

Because time resolution calls for noise-to-slope ratio optimization, while signal-to-noise ratio is the relevant figure for amplitude measurement, the simultaneous acquisition of the two quantities is often performed by splitting the signal after the preamplifier in two independent paths. In one of them, the information is processed by a fast shaper before being presented to the discriminator. As we have seen previously, in order to optimize the slope-to-noise ratio for timing measurement, the shaping time must be possibly matched to the detector collection time. In the other path, the preamplifier signal is fed to a shaper with a longer integration time, which can be chosen to be as long as allowed by the expected event rate or by the detector leakage current. The peak value can then be captured by a sample-and-hold and used for energy measurement as well as for time-walk correction.

A single amplitude measurement is, however, not sufficient to fully solve the time-walk issue in situations, often occurring in practice, where the signal before the discriminator also experiences rise-time variations (Spieler 2006).

These can originate from intrinsic imperfections of the front-end electronics, such as slew-rate limitations. However, even in the case of an ideal readout circuit, the statistical nature of the charge-generation process in the sensor determines rise-time fluctuations, which also occur between signals having the *same* total charge. We have seen previously that, when a particle crosses a semiconductor sensor, it creates along

its path a cloud of electron–hole pairs that, drifting under the electric field, induce signal on the sensor electrodes. However, for the same deposited charge, the ionization density along the particle trajectory will change from one signal to another, thereby introducing a variation in the signal shape and hence in its rise time.

In order to provide immunity also to pulse-shape variations and to push further the timing performance of the system, more-sophisticated techniques for amplitude measurement, like multiple-threshold or waveform sampling, must then be adopted.

Another method for time-walk correction is to employ an analog signal-processing scheme that provides an amplitude-invariant timing signal. The method does not increase the data bandwidth of the system, providing a single accurate piece of information per event. However, it increases the complexity and power consumption of the analog front-end circuitry. In practice, the choice between the two methods is not obvious and depends on many parameters that need to be evaluated carefully on a per-case basis.

## 12.3 CONSTANT-FRACTION DISCRIMINATOR

### 12.3.1 Principle of Operation

One method for accurate time measurements implementing online time-walk compensation is the constant-fraction discriminator (Figure 12.2). The aim of the circuit is to create a variable threshold that tracks the amplitude of the signal, remaining always at a particular fraction of the maximum amplitude of the signal (thus the *constant-fraction* name). The circuit works by comparing a delayed copy of the input signal with an attenuated one. The result is a bipolar signal with the zero-crossing independent of the amplitude of the input signal, eliminating walk error. Ideally, one

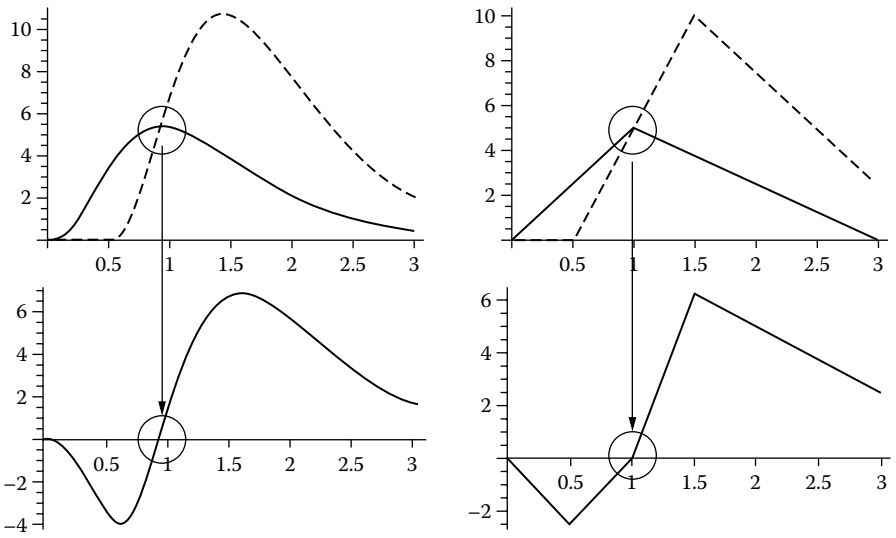



**FIGURE 12.2** Constant-fraction discriminator.

can set the threshold level to (a) optimize the jitter behavior, setting the threshold fraction at the point of the signal shape where the slope is maximum (as discussed earlier), and (b) eliminate the amplitude dependency of the timing signal, both at the same time.

Considering again a triangular signal shape with the leading-edge given by

$$y(t) = A\,t/t_{\mathrm{p}} \tag{12.2}$$

where $A$ is the amplitude and $t_{\mathrm{p}}$ is the peaking time of the signal, the crossing time is given by

$$y\left(t - t_{\mathrm{d}}\right) = fy(t) \tag{12.3}$$

$$A\frac{(t - t_{\mathrm{d}})}{t_{\mathrm{p}}} = fA\frac{t}{t_{\mathrm{p}}} \tag{12.4}$$

$$t = \frac{t_{\mathrm{d}}}{1 - f} \tag{12.5}$$

where $t_{\mathrm{d}}$ and $f$ are the delay and attenuation parameters of the constant-fraction discriminator, respectively. This analysis assumes that the crossing point is always on the linear rising portion of the input signal; otherwise, Equation (12.2) is no longer valid.

Equation (12.5) shows that the zero-crossing time of the constant-fraction bipolar signal is independent from the amplitude of the signal. This independence also continues to hold in the general case, providing that the signal shape is invariant.

As a rule of thumb, one practice is to set the fraction parameter of the CFD to one-half and the delay parameter to half of the full rise time of the signal. In this way, the zero-crossing time matches the peak of the input signal, and the amplitude-tracking threshold is set to 50% of the amplitude of the signal. In some cases, this is a good guess of the optimal condition. Mathematically, for a CR-RC shaping, the slope is maximum at $t = 0$. In practice, the real pulse shape is the result of a more complicated transfer function that includes a multiplicity of poles that come, for instance, from the limited open-loop bandwidth of the amplifier circuits involved, which are usually neglected in first approximation. These additional poles introduce more curvature at the beginning of the pulse (Spieler 2006), pushing the point of maximum slope toward the middle of the pulse shape. The effect is more prominent in very low-power systems, where amplifiers with high gain-bandwidth product cannot be afforded. In a more general case, the condition that the zero-crossing time of the CFD signal matches the peaking time of the input signal, considering again Equation (12.5), implies

$$t_{\mathrm{d}} = (1 - f)t_{\mathrm{p}} \tag{12.6}$$

In this case, the fraction parameter is free and can be set to match the maximum slope of the input signal, for the lowest jitter.

Equation (12.5) may suggest that the crossing time of the bipolar constant-fraction signal does not depend on either the amplitude or the rise-time of the signal. Unfortunately, the immunity to rise-time variations arises from the triangular signal approximation and does not hold anymore for the general case. Variation of the rise time may occur due to distortion effects, which in fact change in a nonlinear way the shape of the signal, introducing a timing error. Another source of signal shape variation is the statistical nature of the signal formation inside the volume of the semiconductor detector.

If the delay parameter of the CFD is set to a value lower than in Equation (12.6), the resulting circuit will also partially compensate a certain variation in the peaking time of the signal. In this case, the effective fractional threshold will move lower for longer peaking times and higher for shorter signals, partially compensating the variation. However, since the crossing point happens before the attenuated signal has reached its peak, the slope of the resulting bipolar signal is degraded, and the overall performance suffers.

Another practice is to set the relative threshold as low as possible above the initial knee of the signal, but sufficiently higher than the noise level. This setting also gives a certain degree of immunity to pulse-shape variations, since large variations in the rising time of the signal are needed to produce significant time error. As in the previous situation, the slope of the bipolar signal is degraded, with negative impact on the performance. Moreover, the resulting bipolar signal is significantly asymmetrical, with a short and weak initial displacement from the baseline, before the zero-crossing point, thus making it difficult for the zero-crossing detection circuitry to operate.

### 12.3.2 Issues in Monolithic CFD Design

Traditionally, the constant-fraction method was used in conjunction with scintillator detectors read out by bulky discrete circuitry, so the delay could be implemented using a coaxial cable of a particular length. Since the front ends for the semiconductor detectors are mainly integrated, the delay line needs to be replaced by another monolithic-compatible implementation of the delay-transfer function.

A particular case of the constant-fraction function is the differentiator that results by setting the fraction parameter to 1 and the delay infinitesimally small. A CR differentiator, for instance, provides a bipolar signal with the zero-crossing time in correspondence with the peak of the signal, requiring no fraction circuit. One drawback of this method is the fact that the overdrive and the slope of the bipolar signal at zero-crossing depend on the trailing edge of the input signal, with additional timing jitter and walk (Jackson et al. 1997).

Other constant-fraction shaping methods were investigated by Jackson et al. (1997), such as a simple RC low-pass filter, a distributed RC delay line, or Nowlin (1992) method. None of these methods can beat the ideal delay-line implementation, the one coming closest to it being the RC delay-line.

For a practical example, we have considered a front-end system that shapes the impulse coming from the detector using a typical CR-RC shaper (see [Spieler 2006] for

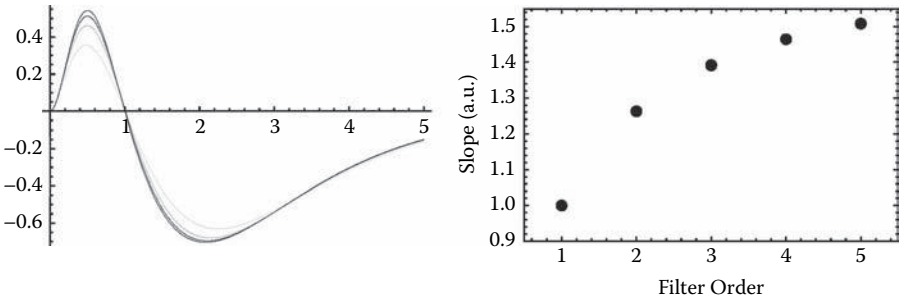



**FIGURE 12.3** CFD signals (left) and the relative slope (right) of the bipolar signal at zero-crossing for different RC-filter orders.

a detailed description) and a constant-fraction discriminator where, instead of the delay line, distributed RC filters of different orders are used. Figure 12.3 shows the bipolar signal and the slope of the bipolar signal at zero-crossing up to the fourth order.

CMOS circuits are famous for rather high offset voltages due to process parameter variation, especially in the case of fast circuits, which are consequently designed to be very small. The presence of offset voltage at the input of the zero-crossing discriminator can destroy the amplitude invariability, creating residual time-walk.

The absolute error introduced by the discriminator offset is equal, in first approximation, to the ratio between the offset voltage and the slope of the bipolar signal, similar to the jitter (see Equation [12.1]). If all nonlinear effects are neglected, the slope of the constant-fraction bipolar signal is proportional to the amplitude of the input signal, so the error becomes proportional to the inverse of the amplitude. Considering a large dynamic range of at least 10:1, the resulting time-walk is approximately equal to the error produced by the minimum signal.

In our practical example, considering a peaking time of the signal of 5 ns and an amplitude range between 100 mV and 1 V, if the offset voltage of the discriminator is just 1 mV, the resulting time-walk is on the order of 30 ps peak to peak, in the ideal CFD case. If a first-order RC filter is used instead of the ideal delay-line, this figure degrades to more than 50 ps, as the bipolar signal becomes slower. Even if this may still be small enough not to influence significantly the final time resolution when the amplitude distribution is also taken into account, 1 mV of offset may be difficult to obtain in fast CMOS circuits. Moreover, for deep submicron technologies, the low power-supply voltage reduces the available dynamic range for the signal, putting additional pressure on the offset-voltage constraint. Static or dynamic offset compensation schemes are often required in these cases.

### 12.3.3 Practical CFD Implementations

A CFD circuit using a distributed polysilicon RC line implemented in a 1.2-μm CMOS technology (Simpson et al. 1996) (see Figure 12.4a) obtained ±150-ps walk and jitter below 150 ps for a 100:1 dynamic range with an area consumption of $200 \times 950\ \mu m^2$ and 15 mW of power consumption.

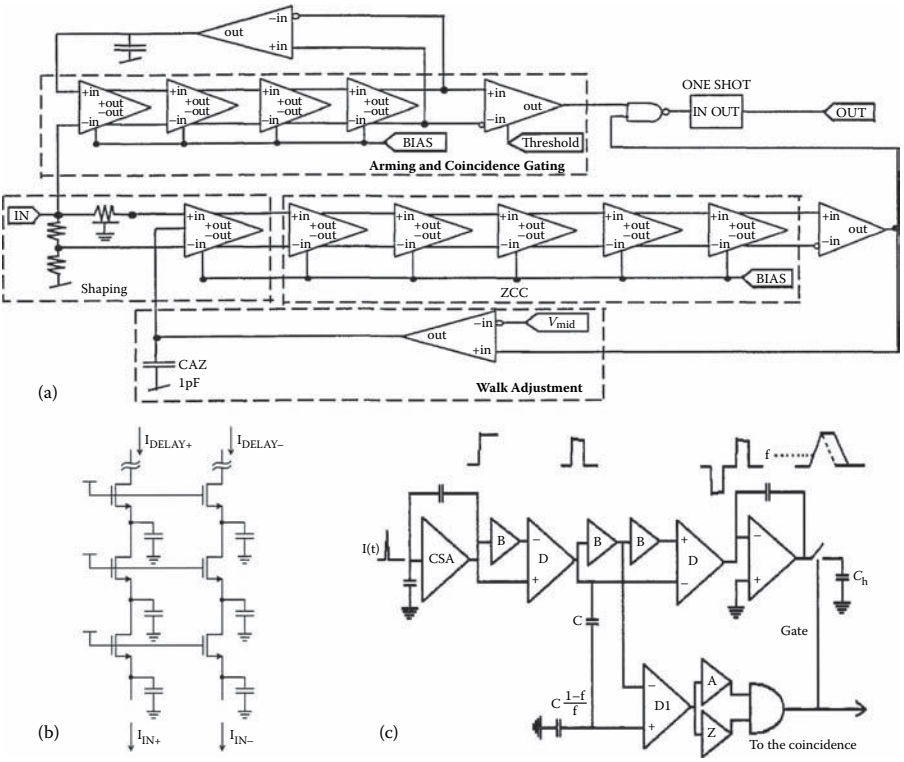



**FIGURE 12.4** (a) Schematic of a CFD monolithic implementation using a distributed RC line-delay element. (From Simpson et al. [1996]. With permission.) (b) Current-mode delay filter example. (From Binkley et al. [2002]. With permission.) (c) CFD circuit using a unity-gain buffer (B) as delay element. (From Manfredi et al. [2001]. With permission.)

Binkley et al. (2002) implemented an integrated current-mode fifth-order low-pass filter using MOS transistors instead of resistors (see Figure 12.4b). While this method provides a good control on the parameters of the delay filter, the DC current supply of the MOS resistors needs to be considerably higher than signal variations for good linearity, thus increasing power consumption. The circuit was implemented in a 0.5-μm CMOS technology and showed intrinsic timing jitter and walk below 100 ps for a 3.7:1 dynamic range, with the expense of 50-mW power consumption.

Manfredi et al. (2001) used a feedback unity-gain buffer as a delay element in the CFD design (Figure 12.4c). The results for a front-end circuit optimized for detector capacitance below 200 fF were 1-ns time-walk and ≈250-ps time jitter, for input charge signals between $2 \times 10^4$ and $5 \times 10^5$ electrons. The power consumption of the circuit was about 3 mW.

In the first cited work, the offset of the zero-crossing discriminator is compensated by a dedicated continuous-time DC feedback loop that controls internally the first stage of the discriminator. In Binkley et al. (2002), a more complicated baseline restorer circuit is used in order to accommodate for higher event rates. In this case, a

current generated by the compensation circuitry is added into the signal path, at the input of the discriminator.

More recently, Engel et al. (2006) designed a 16-channel front-end chip for solid-state detectors, fabricated in a 0.5-μm CMOS process. Each channel of the chip contains (a) a linear branch dedicated to amplitude measurement and peak detection and (b) a timing branch consisting of a constant-fraction discriminator that uses a passive Nowlin circuit and a time-to-amplitude converter. The offset of the zero-crossing discriminator is canceled by an internal dynamic compensation circuit. The chip achieves 1-ns time resolution FWHM and an experimental peak-to-peak time-walk variation below ±1 ns for a dynamic range of 0.5 to 50 MeV, consuming 20 mW in the timing branch.

In the framework of the future NA62 experiment at CERN, Switzerland, R&D work was carried out by the authors to provide a front-end electronics solution to a hybrid silicon pixel detector with 200-ps time resolution for minimum-ionizing particles. The pixel size in this case was of $300 \times 300$ μm, while the detector thickness was chosen to be 200 μm as a compromise between signal strength and minimum material budget. More details on the detector specifications and the experiment itself can be found in Fiorini et al. (2007).

Given the rather large pixel size, which also implies a relatively generous maximum power budget of 1.8 mW per pixel, the authors decided to explore the possibility of implementing a complete pixel front-end based on a constant-fraction discriminator, integrating local time-to-amplitude converter, analog derandomization of the time information, and digitization.

One prototype of the CFD front-end stage was realized in a commercial 0.13-μm feature size CMOS process and reported by Martoiu et al. (2008). In this implementation, the delay and attenuate transfer functions required for the constant-fraction are synthesized by a passive RC differential filter, as shown in Figure 12.5. The four distributed low-pass RC filters simulate a delay line, while the two resistors cross-coupled between the input and the output of the filter form a differential attenuator. One should observe that, by disconnecting the capacitors, the circuit turns into a differential resistive-fraction circuit with inverted output. The ratio between the resistors determines the fraction parameter of the constant-fraction transfer function, and the RC product determines the delay. The passive elements are implemented using polysilicon resistors and vertical plate capacitors provided by the CMOS process.

The output signal of the constant-fraction filter is amplified to near CMOS level by a sequence of fast differential amplifiers. The offset sources up to this stage are mitigated

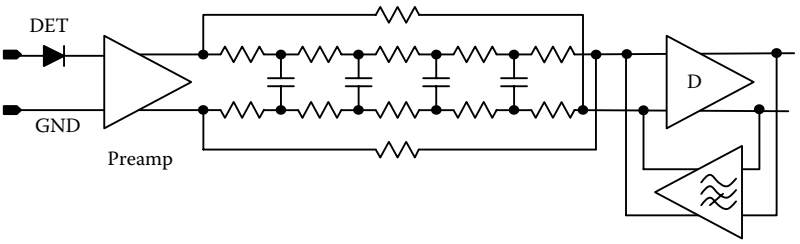



**FIGURE 12.5** CFD-based silicon pixel front end designed for the Gigatracker detector of the NA62 experiment at CERN. (Reported in Martoiu et al. [2008]. With permission.)

by a slow current-mode feedback loop. A transimpedance amplifier is used to inject current into the resistive ladder of the CFD passive filter. The differential component of this current is modulated by the DC error sensed at the output of the discriminator. The slow feedback obtained compensates the relatively high offset of the fast differential amplifiers, as well as the offset sources from the front-end stages. It also provides a high-pass overall transfer function, reducing the low-frequency white and 1/*f* noise.

Jitter and time-walk performance are shown in Figure 12.6 for an input charge pulse between 1 fC and 10 fC. The time-walk curve shows a residual leading-edge behavior at lower amplitudes, while at higher amplitude range, the curve turns back to higher delays. This behavior was found to be caused by the distortion of the signal shape, which takes place in the preamplifier stage. The circuit shows typically less than 80-ps jitter error for the average MIP signal (2.2 fC or 14,000 electrons, in this case) and less than 270-ps peak-to-peak time-walk error for the whole amplitude range.

Weighting the measured results by the expected Landau distribution of the detector charge yield, the combined time resolution is found to be on the order of 100 ps (Figure 12.7). The amplitude distribution is concentrated in the central part of the

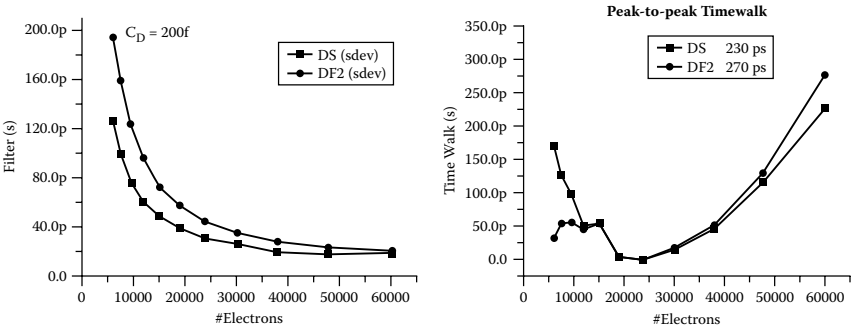



**FIGURE 12.6** Jitter and time-walk measurements.

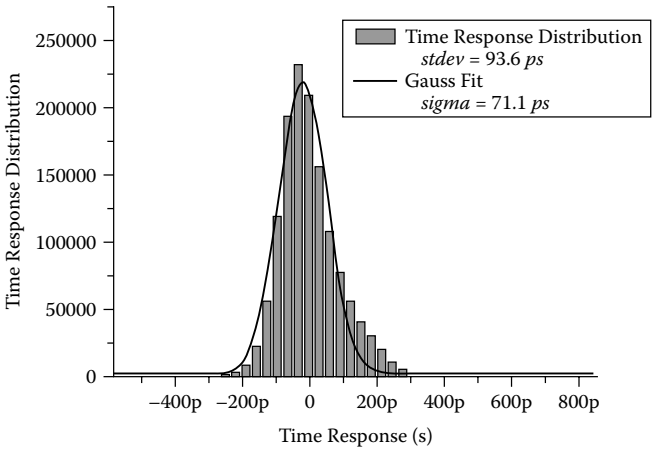



**FIGURE 12.7** Time-resolution accounting for the expected Landau charge-yield distribution.

dynamic range, around the average 2.2 fC (14,000 electrons) yield. Since the variation of the time-walk in this region is small, its contribution to the total timing performance is close to negligible.

## 12.4 TIME-WALK CORRECTION WITH AMPLITUDE INFORMATION

### 12.4.1 Time-over-Threshold: Basic Principle and Practical Implementations

It was shown in Section 12.2 that a simple leading-edge discriminator suffers from time-walk due to different threshold-crossing times experienced by input signals with different amplitude. Hence, the parallel measurement of the amplitude provides a means to compensate this effect off-line, back-projecting the experimental time response at the correct position. One needs to know quantitatively the correspondence between the signal amplitude and the time-walk error, and then subtract it from the measured leading-edge time affected by the walk error. All uncertainties in the amplitude measurement and time-walk to amplitude mapping add up to adversely affect the timing precision. Here the relevant error sources are, for example, the noise in the amplitude measurement, the ADC resolution, as well as the accuracy of the parameterization of the time-walk as a function of the signal amplitude. In principle, the best opportunities for off-line corrections are provided when the full signal waveform is digitized.

In a recent paper by Ruckman and Varner (2009), a switched capacitor array with a sampling frequency of up to 6 GHz was employed to capture fast signals delivered by multichannel plate (MCP) photomultipliers. After the off-line software corrections, a final resolution of less than 10 ps rms was demonstrated. In many applications, however, due to the complex circuitry and the power consumption required, full waveform digitization can not be afforded. As we can observe in Figure 12.2, the time spent by a signal above a given threshold depends in general on the signal amplitude. This property can be exploited to perform an amplitude measurement with a minimal circuit overhead.

Known as time-over-threshold (ToT), this approach is particularly suitable for a low-power, multichannel readout system and, owing to its simplicity, it has been adopted in a number of designs. ToT information can, in fact, be easily extracted from a simple binary front end, in which a discriminator follows the input amplifier, by measuring the leading and trailing edges of the comparator against a time reference. Depending on the precision needed, the time measurement can be provided either by counting clock pulses or by using a high-resolution time-to-digital converter (TDC).

A well-detailed example of a front-end circuit exploiting the ToT technique is found in the AToM chip, developed for the silicon tracker of the BaBar experiment at the SLAC National Accelerator Laboratory at Menlo Park, California (Kipnis et al. 1997). In each of the 128 channels, a leading-edge comparator senses the output of a semi-Gaussian shaper. The discriminator output is sampled at 15 MHz, and the resulting bits are stored in a 193-cell-deep digital pipeline. When a trigger signal is issued, a digital processor looks in the bit string for 0-to-1 and 1-to-0 transitions that identify a pulse exceeding the threshold. The number of 1's in each pattern gives the

ToT information. In the case of a semi-Gaussian shaper, the relationship between the pulse duration and the signal amplitude follows a quasi-logarithmic curve, which can also be exploited to perform a signal compression and expand the dynamic range. Interestingly, for large signals, slew-rate limitations in the shaper may cause a distortion of the pulse shape that leads to a more linear ToT-vs.-amplitude curve, a fact that was also observed in the AToM chip. In this ASIC (application-specific integrated circuit), the primary purpose of the ToT feature was to improve the spatial resolution in case the charge is shared among two adjacent strips.

An example of the use of the ToT method for time-walk correction is instead found in the NINO ASIC, developed for the time-of-flight (TOF) system of the ALICE experiment at CERN (Anghinolfi et al. 2004). Implemented in a 0.25-μm CMOS process, the circuit consists of a common-gate input stage, followed by a chain of low-gain and fast differential amplifiers. The output of the discriminator is converted to an LVDS (low-voltage differential signaling) and sent out of chip, where a high-resolution TDC with 25-ps bins is employed to measure the leading and the trailing edge. Also, for the NINO chip, the relationship between the signal charge and the pulse width has a nonlinear behavior, and the signal compression is more pronounced for smaller signals. Exploiting the ToT information, the time-walk error is corrected, and a final time resolution of 50 ps rms is obtained when reading out multigap resistive plate chambers. The power consumption of the circuit is 30 mW per channel, which in this particular application is motivated by the very demanding performance required.

One of the interesting aspects of the ToT approach is that it has the potential of being implemented with very low power consumption. Therefore, it is the natural option if one wants to equip highly segmented radiation sensors, like silicon hybrid pixel detectors, with energy and time-resolution capability. The front-end chip developed for the ATLAS pixel system (Ivan Peric et al. 2006) is one interesting example.

More recently, a versatile pixel detector front-end, called Timepix, has been reported (Llopart et al. 2007). The ASIC contains a matrix of $256 \times 256$ independent cells. In an area of $55 \times 55$ μm, each cell incorporates a front-end amplifier, a discriminator, and a programmable counter that allows three different operating modes. In the "event counting mode," the counter is incremented by one each time the comparator fires, a modality that is suitable, for instance, for imaging applications. In the "ToT mode," the counter is driven by a reference clock, which is provided by an external source and distributed to the whole matrix. The counter is enabled while the comparator signal is above the threshold level, thus allowing the measurement of the deposited energy. Finally, in the arrival-time mode, the counter is triggered by the comparator transition and frozen by a common stop signal. An interesting feature of the Timepix chip, which is fabricated in a 0.25-μm CMOS process, is its low power consumption, which is only 12 μW per cell.*

It must be observed that the three modes of operation described here are mutually exclusive, so in the timing mode the ToT information is not available. However,

* Such a low power consumption is of course possible because the very low capacitance typical of pixel detectors allows a good signal-to-noise ratio, even with a small bias current.

each cell can be configured independently. As suggested by Llopart et al. (2007), the small pixel size in some applications allows exploiting the benefit of charge sharing to perform both measurements simultaneously. In such a case, the chip could be configured in a chessboard pattern, with one cell working in ToT mode and the neighbor one in timing mode. This would make available for the same hit the full information necessary for the correction of the time-walk. In practice, the application of a correction procedure entails several issues, which are discussed in the practical example in the following subsection.

### 12.4.2 Time Resolution of a Low-Power ToT Cell in 0.13-μm CMOS: A Case Study

To investigate the performance of the ToT approach for time-walk correction in the case of low-power circuits suitable for multichannel semiconductor radiation sensors, the time resolution of a pixel front-end cell implemented in a 0.13-μm CMOS process was measured (Martoiu 2007). The chip used as a test vehicle was developed in the framework of an R&D project for the front-end electronics of the microvertex detector of the PANDA experiment, planned at the future FAIR facility at GSI. The circuit is designed to provide simultaneously time-stamping of the events and energy information.

A block diagram of the circuit is shown on the left part of Figure 12.8, while the right-hand side shows simplified transistor-level schematics of the input amplifier and of the discriminator. The circuits are kept as simple as possible to minimize power consumption and area occupation. The detector current is integrated on the feedback capacitor $C_f$, which has a value of 10 fF. The capacitor is discharged with a constant current, providing an output waveform with a triangular shape. A low-bandwidth transconductor compensates the sensor leakage current and allows locking the DC voltage at the preamplifier output to a reference voltage. A calibration capacitor is used to inject artificial signal. The preamplifier is implemented with a single-ended telescopic cascode. The maximum applicable power supply is only 1.5 V, but the two nMOS transistors operate deeply in weak inversion, so they need only 100 mV to stay in saturation. The current source works in strong inversion and with a transconductance as low as possible to minimize its contribution to the noise. During the measurements, the circuit was powered with a single-rail power supply of 1.2 V, while the total power consumption was 10 μW. With these settings, a linear dynamic range of 0.7 V was obtained at the preamplifier output. It is well known, however, that such an architecture can provide a linear measurement of the injected charge, even when the amplifier saturates, as long as the current source providing the feedback current works properly. In the tests, the signals were fed directly to the calibration capacitor, while the comparator output was captured with a fast digital oscilloscope.

The measured time-walk curve as a function of the time-over-threshold is shown in Figure 12.9 for three different channels of the same chip. The time error was measured directly versus the time-over-threshold, as opposed to the amplitude level of the input signal. This was done to avoid the propagation of the parameterization

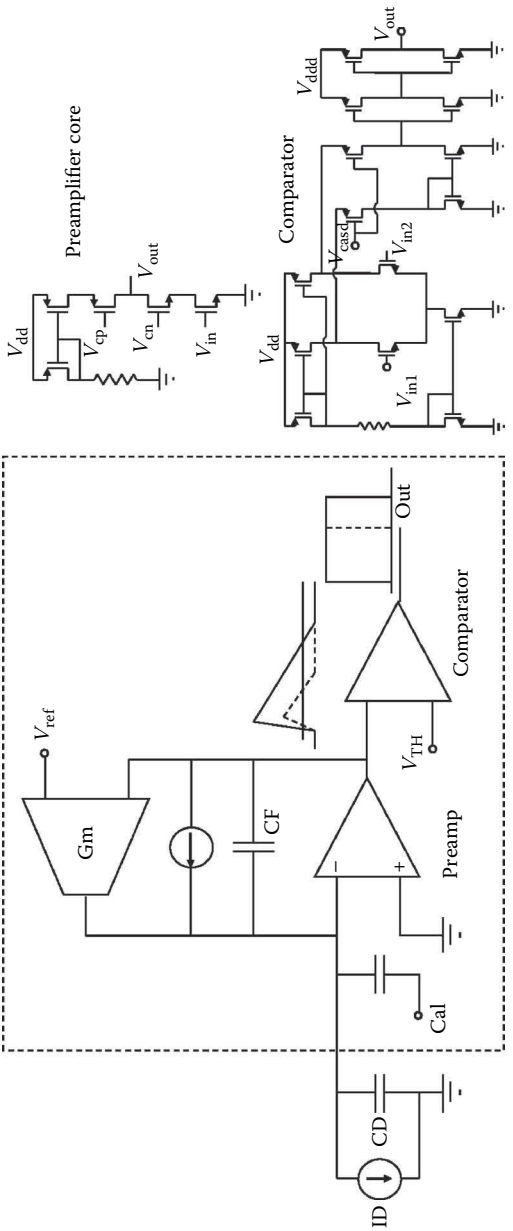



**FIGURE 12.8** Block diagram and simplified schematics of the front-end prototype for the Microvertex detector of the PANDA experiment at FAIR, GSI.

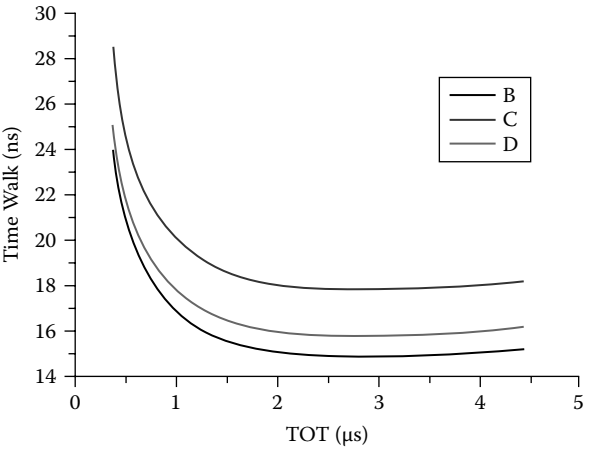



**FIGURE 12.9** Time-walk versus ToT without correction.

errors between time-over-threshold and amplitude on the one hand, and between amplitude and time-walk, on the other. The measurements were performed with input amplitudes ranging from 1.4 fC to 22 fC in 25 logarithmically spaced steps.

Ideally, the signal shape at the output of the preamplifier should be a perfect triangle. In this simple case, the equation of the leading edge would be written as

$$y(t) = \frac{A}{\tau_p} t \tag{12.7}$$

where $A$ is the amplitude, $\tau_p$ is the peaking time, and $t$ the time variable.

The running time-walk can be defined as the time the signal needs to reach the threshold voltage $y(t_w) = V_{thr}$

$$t_w = \frac{V_{thr}}{A} \tau_p \tag{12.8}$$

It follows that, in the ideal case, the running time-walk curve is proportional to the inverse of the amplitude of the preamplifier output signal.

In reality, the leading edge of the signal is not linear, its form determined by the shaping effect of the preamplifier while still in a linear range, and the large-signal nonlinearity of the circuit. For a more detailed study, one should also include the errors induced by the discriminator, i.e., the slew-rate dependence of the propagation time. However, a full analytical model would bring too much complexity, so a simpler fitting approach was preferred. The time-walk curve was parameterized using a third-order polynomial function of the inverse of the ToT measure

$$t_w = c_0 + c_1 \frac{1}{\mathrm{ToT}} + c_2 \left( \frac{1}{\mathrm{ToT}} \right)^2 + c_3 \left( \frac{1}{\mathrm{ToT}} \right)^3 \tag{12.9}$$

Figure 12.10 shows the residual of the time-walk correction. Using 25 measured points to calculate the parameters in Equation (12.9), the time error was reduced from as much as 10 ns down to 300–400 ps peak to peak. Using only five points for the interpolation of the time-walk versus ToT curve, the residual error increases by approximately 50%. This analysis takes into account the fact that the calibration of a high number of pixels is impractical if too many calibration points are considered. The result stresses the importance of the accuracy of the parameterization in this analysis.

Another important component of the time resolution is the jitter, which as we have seen in Section 12.2 is defined as the ratio between the rms noise level at the input of the discriminator and the signal slope at the threshold-crossing point. Because the leading edge is used to extract the primary timing information, one might surmise that the jitter on the leading edge should provide the only important contribution. As a matter of fact, since the time information is corrected for time-walk errors using the amplitude information, the random error on the amplitude measurement (which is in nature equivalent to the jitter) must also be taken into account. Moreover, in a system that uses time-over-threshold, the error on the amplitude is related to the jitter of both the leading edge and the trailing edge of the signal, which defines the ToT. Figure 12.11 shows the standard deviation of the ToT measurement for the practical example considered here. The use of a constant current source in the feedback path to linearize the amplitude measurement has the disadvantage of a high jitter of the trailing edge, which can reach 10% for low-amplitude signals. As shown in Figure 12.11(b), its contribution to the overall jitter-induced error is very important for the lower part of the amplitude range, while it is negligible at high amplitudes.

Figure 12.12 shows the standard deviation of the leading-edge time (Figure 12.12a) and the square root sum of the latter with the contribution of the time-walk correction. An important point here is that the amplitude-driven time-walk correction comes at the expense of a deterioration of the original leading-edge jitter at low amplitudes, which in this case jumps from 300 ps to 1 ns.

Two important remarks should finally be made. First, the main reason the power consumption is so low is that, as for the Timepix chip discussed previously, the circuit is optimized to work with a detector with very low capacitance (typically 200 fF).

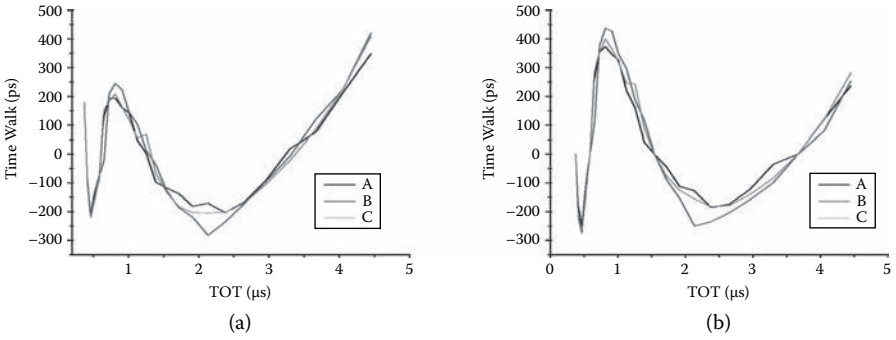



**FIGURE 12.10** Time-walk with amplitude correction using a third-order polynomial function. (a) 24 calibration points; (b) 5 calibration points.

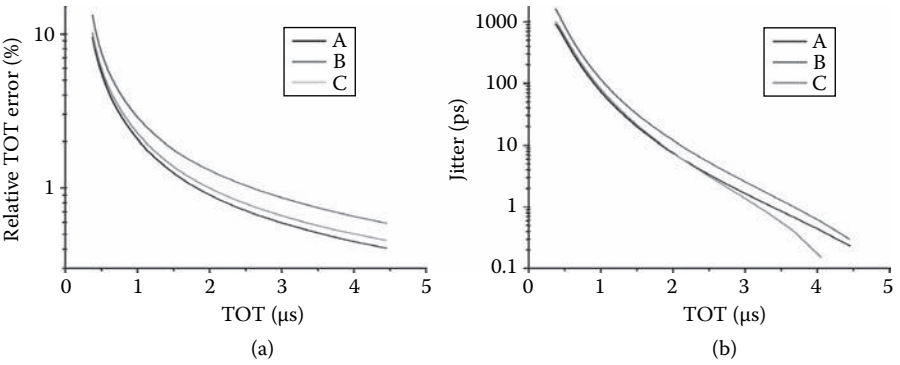



**FIGURE 12.11** (a) Measured relative TOT jitter. (b) Contribution of the time-walk correction to jitter.

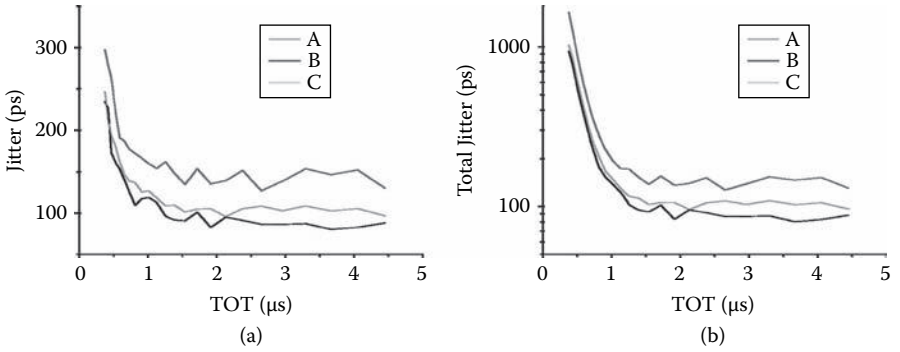



**FIGURE 12.12** (a) Measured jitter of the leading-edge discriminator transition. (b) Combined jitter.

Therefore, the input amplifier can be biased with a small current (on the order of 1 μA) while maintaining an adequate noise level. Second, the resolution obtained here is the one of the electronics alone, so it represents a best-case approximation. When designing a real system, one should not overlook all the aspects related to the sensor behavior, such as rise-time signal variations for the same deposited charge, which contribute significantly to the final timing accuracy.

## REFERENCES

Anghinolfi, F., et al. 2004. NINO: An ultra-fast and low-power front-end amplifier/discriminator ASIC designed for the multigap resistive plate chamber. *Nucl. Instr. Meth. A* 533:183–187.

Bichsel, Hans. 1988. Straggling in thin silicon detectors. *Rev. Mod. Phys*. 60:663.

Binkley, D. M., et al. 2002. A 10-Mcps, 0.5-μm CMOS constant-fraction discriminator having built-in pulse tail cancellation. *IEEE Trans. Nucl. Sci*. 49:1130–1140.

Engel, G. L., et al. 2007. A multi-channel integrated circuit for use in low and intermediate-energy nuclear physics—HINP16C. *Nucl. Instr. Meth. A* 573:418–426.

Fiorini, M., et al. 2007. The P326 (NA48/3) Gigatracker: Requirements and design concept. *Nucl. Instr. Meth. A* 572:290–291.

Green, D. 2000. The physics of particle detectors. In *Cambridge monographs on particle physics, nuclear physics and cosmology*, 39–40. Cambridge: Cambridge University Press.

Jackson, R. G., et al. 1997. Integrated constant-fraction discriminator shaping techniques for the PHENIX lead-scintillator calorimeter. *IEEE Trans. Nucl. Sci*. 44:303–307.

Kipnis, I., et al. 1997. A time-over-threshold machine: The readout integrated circuit for the *BABAR* silicon vertex tracker. *IEEE Trans. Nucl. Sci*. 44:289–297.

Knoll, G. F. 2000. *Radiation detection and measurement*. 3rd ed. New York: John Wiley & Sons, 659–679.

Landau, L. D. 1944. On the energy loss of fast particles by ionization. *J. Phys. USSR* 8:201.

Leo, W. R. 1987. *Techniques for nuclear and particle physics experiments*. Heidelberg: Springer-Verlag, 47–50.

Llopart, X., et al. 2007. Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurement. *Nucl. Instr. Meth. A* 581:485–494.

Manfredi, P. F., et al. 2001. Acquisition and processing of multiparametric information from a pixel matrix. *IEEE Trans. Nucl. Sci*. 48:1244–1248.

Martoiu, S. 2007. Design, test and system integration of front-end electronics for particle detection in high energy and nuclear physics experiments. PhD thesis in experimental physics, University of Turin, Italy.

Martoiu, S., et al. 2008. A low power front-end prototype for silicon pixel detectors with 100-ps time resolution. In *Nuclear Science Symposium Conference Record*, *2008. NSS '08. IEEE*, 2958–2961.

Moses, W. W. 2007. Recent advances and future advances in time-of-flight PET. *Nucl. Instr. Meth. A* 580:919–924.

Nowlin, C. H. 1992. Low-noise lumped-element timing filters with rise-time invariant crossover times. *Rev. Sci. Instr*. 63:2322–2326.

Peric, Ivan, et al. 2006. The FE13 readout chip for the ATLAS pixel detector. *Nucl. Instr. Meth. A* 565:178–187.

Ruckman, L. L., and G. S. Varner. 2009. Sub-10 ps monolithic and low-power detector readout. *Nucl. Instr. Meth. A* 602:438–445.

Simpson, M. L., et al. 1996. A monolithic, constant-fraction discriminator using distributed R-C delay line shaping. *IEEE Trans. Nucl. Sci*. 43:1695–1699.

Spieler, H. 1982. Fast timing methods for semiconductor detectors. *IEEE Trans. Nucl. Sci*. 29:1142–1158.

Spieler, H. 2006. *Semiconductor detector system*. Oxford: Oxford Science Publications, 179–188.

# 13 Time-to-Digital Converter Circuits in Radiation Detection Systems

*Sachin Junnarkar*

## CONTENTS



## 13.1 INTRODUCTION

Timing measurement, a fascinating instrumentation problem, has long challenged researchers. Great efforts have been dedicated to ensure precise timekeeping and measurements of time both for commonplace applications, such as wristwatches, and complex ones, as in many space-science instruments, such as particle physics systems,

X-ray and UV imagers, and laser altimeters. The latter require sharply exact measurements of time intervals at low power, high speed, and fine resolution [1–5]. In nuclear-particle physics experiments, drift-based tracking detectors are used to obtain highly accurate information about the energy, momentum, time of occurrence, and/or position of incidence of a particle on the detector. For a majority of these applications, the instruments consist of radiation detectors wherein, in the simplest case, charge is induced on a set of two electrodes. Only a fraction of this charge is available for measurement, and it is collected and amplified by an electronics preamplifier [6].

To extract timing information from the collected charge, the electronics chain usually consists of the following components: shaping amplifier and filter followed by a constant-fraction discriminator (CFD), then a timing measurement circuitry or an analog-to-digital converter (ADC), and finally the digital signal-processing circuitry [7]. While such systems give better timing resolution than the earlier analog CFD techniques, the large number of ADCs required—which may be as many as the number of front-end channels (which usually range from 100 to more than 100,000)—dominate the cost and power requirements of the system. In some systems, the shaping amplifier and filter is followed by a zero-cross discriminator (ZCD) [8, 9] rather than a CFD. An analog CFD or ZCD output is used as the measure of time. Two kinds of mechanisms are employed to digitize the CFD/ZCD output. A time-to-amplitude converter (TAC) measures the interval between two CFD/ZCD pulses—start and stop—and generates an analog output pulse proportional to the measured time difference between them. This analog output may be digitized using a conventional analog-to-digital converter (ADC), thus giving a digital representation of time difference between the start and stop pulses. A time-to-digital-converter (TDC) is a device for converting these CFD/ZCD output signals directly into a digital representation of their absolute or relative time of occurrence (Figure 13.1).

Although TACs are very advanced timing-measurement circuits, they require ADCs. Furthermore, as the channel counts increase for a given detector, TACs become expensive and high-power-consuming solutions. TDCs, on the other hand, can be designed

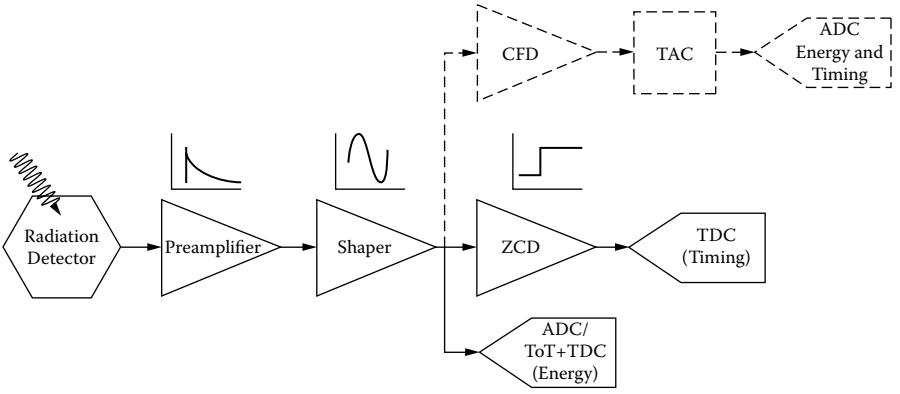



**FIGURE 13.1** Radiation detector signal chain showing conventional signal chain in dashed boxes versus contemporary processing units shown in solid. (From Junnarkar et al. [10]. © IEEE 2009/2010. With permission.)

to be very compact, with multiple channels packed together as a part of a module, with much lower power consumption while retaining very high precision. As many as 24 TDC channels have been realized in a single FPGA (field-programmable gate array) device with subnanosecond resolution [8]. Only a few percent of the total system power budget is used for those TDC channels, because the same FPGA is serves as a workhorse of the data acquisition system. A custom application-specific integrated circuit (ASIC)-based solution also provides similar form factor, low power dissipation, and performance. In this chapter, we visit both implementations of the TDC.

## 13.2 TDC APPLICATIONS

TDCs are needed extensively in high-energy physics experiments. In drift-based tracking detectors, a time resolution on the order of 1 ns normally is sufficient. However, in time of flight (TOF) detectors, a time resolution of a few tens of picoseconds is often required [11–15]. Some commercial PET scanners employ a TDC to measure time accurately [16, 17]. Recently, TDCs have been applied to frequency synthesis in delay-locked loops (DLLs) for faster acquisition of data and to avoid false locking. With the advent of digitally intensive and all-digital fractional phase-locked loops (PLLs) in deep-submicrometer complementary metal-oxide-semiconductor (CMOS) scaling, the TDC is becoming an attractive replacement for the conventional phase/frequency detector and charge pump [18]. TDCs also have been employed as low-power, highly accurate, portable temperature sensors [19]. The key building blocks in wireless transceivers incorporate TDCs for realizing digital frequency synthesizers [20–22]. As gigahertz (GHz) speeds become more common in high-speed data communications and microprocessors, the jitter requirements for PLLs have fallen into the range of picoseconds. TDCs are used for real-time jitter measurements in such applications [23].

Presently, there are three experiments at Brookhaven National Laboratory for which we must resolve the problem of time-to-digital conversion. Two experiments involve positron emission tomography (PET)-based scanners, and the third is a time-of-flight (TOF) type of experiment based on NaI(Tl) scintillators and a photomultiplier tube (PMT) detector chain. Given the expanse of TDC applications, we describe these three experiments in the following subsections.

### 13.2.1 Positron Emission Tomography

Positron emission tomography (PET) is a nuclear medicine imaging method for obtaining three-dimensional images of physiological processes in the body. In studies with PET imaging systems, a positron-emitting radionuclide, the tracer, is injected intravenously into the patient and is taken up by active molecules. The radionuclides typically employed are short-lived ones, viz., carbon-11, nitrogen-13, fluorine-18, and oxygen-15, with lifetimes ranging from 2 to 110 minutes. When a positron comes into contact with an electron within the body, the two particles annihilate and produce two time-coincident 511-keV gamma rays that are emitted approximately 180 degrees apart [16]. These collinear gamma rays are detected by gamma-ray detectors, such as scintillation detectors, e.g., lutetium oxyorthosilicate

(LSO) and lanthanum bromide ($LaBr_3$). In the PET scanner, the many gamma-ray detectors are arranged in parallel rings that encircle the patient being imaged.

A time measurement is required to detect the coincidence of two opposing gamma rays that hit two detector pairs. Photodetectors, such as photomultiplier tubes (PMTs) or avalanche photodiodes (APDs) convert the scintillated photons and generate an electric charge. The front-end electronics chain—preferably consisting of a preamplifier, shaper, constant-fraction discriminator similar to one shown in Figure 13.1, and a timing-measurement circuit—extracts from the collected charge information such as energy, time of occurrence, and position of incidence. Time of occurrence of the events is the quantity of utmost importance for reconstructing the images. Online or offline coincidence processing systems plot histograms of the recorded timing information from the detected colinear annihilation events of interest. The complexity and large channel-count in a typical PET system necessitates developing custom and integrated front-end electronics using standard, low-cost, and readily available CMOS- or FPGA-based technologies with low power consumption.

Besides the medical imaging of patients, PET finds applications in fields ranging from in vivo high-resolution imaging of neurophysiologic activity in animals, wherein the rat conscious animal PET (RatCAP) [24] and wrist detector are examples of two specific applications to measurements of arterial-input function in humans. The front-end electronics signal chain for these detectors consists of a preamplifier-shaper followed by a ZCD and TDC [25].

A significant amount of research with PET has focused on small-animal studies. In studying abnormalities, such as addiction, hyperactivity, and stress, it is important to correlate animal behavioral patterns with PET information from humans [26]. The architecture of many contemporary PET scanners is large, as these systems rely on highly segmented arrays of scintillating crystals coupled with photomultiplier tubes. With a few exceptions, such as in works by Momosaki et al. [27] and Tsukada et al. [28], the animals must be anesthetized, since any movement will severely distort the quality of the images obtained. However, anesthesia not only constrains the ability to relate behavioral patterns and PET information, but it also suppresses many important neurological activities that are being studied. A compact electronics architecture resolves the mobility issue [9].

### 13.2.2 Associated Particle Technique, Time-of-Flight Experiments

In the associated-particle technique (APT) [29], the alpha particle associated with a 14-MeV [30, 31] neutron produced by the $^3H(d,n)^4He$ reaction in the neutron generator is emitted in nearly the opposite direction. When a neutron collides with the soil matrix, one probable interaction is an inelastic scattering by a C atom, producing the characteristic gamma-ray emission. The elapsed time between detection of the alpha particle and detection of the gamma ray reveals the distance traveled by the neutron and provides the depth of neutron interaction with C in the soil, as shown in Figure 13.2. Since 14-MeV neutrons travel at a speed of 5 cm/ns, it is important to achieve subnanosecond time resolutions to be able to resolve depth profiles of C layers less than 5-cm thick [32].

Fast-rising PMT timing signals can be AC-coupled and directly read by the FPGA inputs with little or no signal conditioning (Figure 13.3). Modern-day high-speed

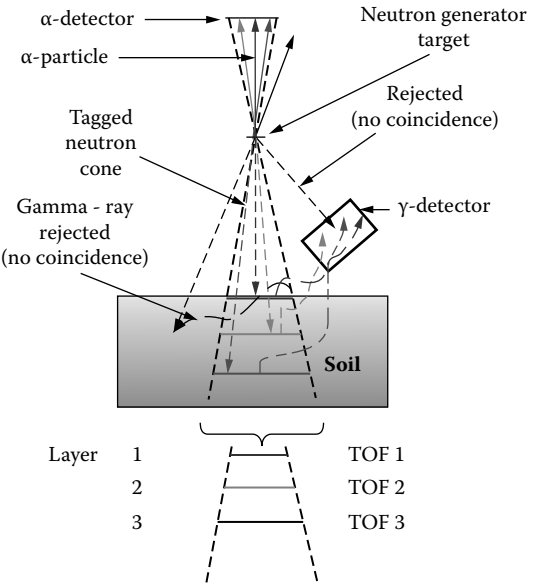



**FIGURE 13.2** Associated particle technique, showing the directional correlation between the alpha particle and a 14-MeV neutron. Layers 1, 2, and 3 should conceptually produce three different neutron time-of-flight signatures.

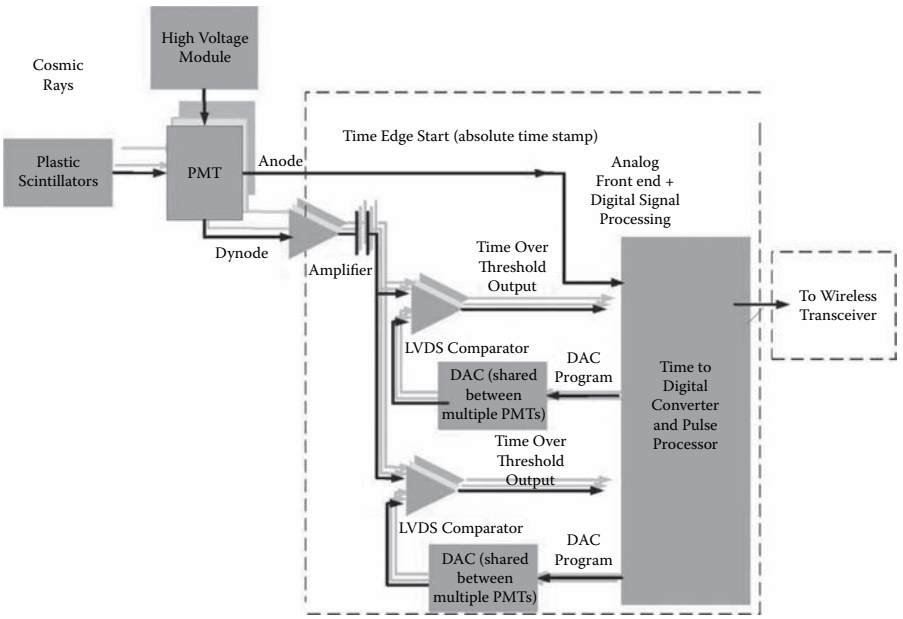



**FIGURE 13.3** Block diagram of front-end signal processing and digital data acquisition for MARIACHI project.

field-programmable gate array (FPGA) devices can be exploited for making high-resolution timing measurements. An FPGA-based time-to-digital converter using a single tapped delay line (TDL) with subnanosecond timing resolution has been successfully realized [9]. Digitizing the time of occurrences of PMT output pulses using a TDC, realized using the TDL technique, should, in principle, provide the timing of the first photoelectron with error limited by TDC resolution, very negligible PMT jitter, and photostatistic noise. To obtain reasonable energy discrimination for elemental analysis, a digital pipeline of ADC samples is a good solution.

### 13.2.3 MARIACHI Experiment

The MARIACHI project has had great success at the Brookhaven National Laboratory on Long Island, New York. The MARIACHI philosophy of "learning by doing" holds great potential for its implementation on a larger scale. Continental-scale coverage requires upgrades to existing MARIACHI instrumentation. Next-generation MARIACHI detectors require low-power, battery-operated FPGA-based front-end instrumentation with a wireless data-acquisition system.

A novel field-programmable gate array (FPGA)-based electronics architecture for picoseconds time-of-flight spectroscopy can be used for detection of cosmic rays. The detector includes plastic scintillators coupled to a photomultiplier tube.

Fast-rising PMT timing signals can be directly read by the FPGA inputs without any signal conditioning (Figure 13.2). Modern-day high-speed FPGA devices can be exploited to make high-resolution timing measurements. Digitizing the time of occurrences of PMT output pulses using TDC should provide the timing of the first photoelectron with error limited by TDC resolution, very negligible PMT jitter, and photostatistic noise. To obtain reasonable energy discrimination, a TDC quantizer will be used as an ADC using the time-over-threshold technique. The fundamental resolution limit on an ADC using a TDC was found to be <5 bits. This resolution can be increased beyond the fundamental limit by sampling the waveform at multiple thresholds. In order to cover the complete PMT dynamic range and obtain enough samples to allow reconstruction, thresholds will be logarithmically spaced. The trade-off in this approach would be to arrive at the optimum number of thresholds to achieve sufficient ADC resolution to perform energy discrimination vs. the power consumption in the LVDS (low-voltage differential signaling) receivers, TDC circuits, and DACs (digital-to-analog converters). Synchronizing multiple detector TDC outputs will be a challenge. State-of-the-art global-positioning-system receiver clocks suffer from ±1-ns uncertainty.

## 13.3 TDC CHARACTERISTICS

A TDC quantizer is similar to the very well-studied and commonly known ADC quantizer. The quality of a TDC is judged by similar parameters as an ADC, with high resolution being the foremost of all the TDC's specifications. Definitions of important terms are as follows:

*Resolution*: The smallest time-interval that can be digitized in a TDC.

*Dynamic range of phase detection*: The maximum phase error detectable or maximum measurable time interval using the TDC circuit.

*Robustness*: TDC robustness is specified over a range of variations in process, voltage, and temperature (PVT).
*Dead time*: Time interval required to complete the digitization process or the minimum delay between two successive digitization operations is referred to as dead time.
*DNL (differential nonlinearity) error*: The difference between a real least-significant bit (LSB) and an ideal LSB.
*INL (integral nonlinearity) error*: The difference between a real transition point and an ideal transition point.
*FoM (figure of merit)*: The incremental energy expended in conversion of one of the possible $2^{ENOB}$ quantization steps to the next quantization step, with resolution equal to quantization step size $q$, during conversion of an input time to its digital representation.
*ENOB*: Effective number of bits.

## 13.4 PROCESS, VOLTAGE, AND TEMPERATURE VARIATIONS

Although the previously discussed techniques apparently are straightforward and very simple to implement, in reality delay lines are very susceptible to variations in process, voltage, and temperature, thus complicating life for the designer. Ways must be devised to compensate for, and sometimes to overcome, these inherent nuisances. The following section details the problems associated with these techniques and the solutions developed so far.

### 13.4.1 Temperature Variations

Studying the standard behavior of the MOS transistor can shed light on the effect of temperature on logic gates and delay buffers. The conduction current of the MOS transistor in the saturation region is denoted by

$$I_D = \frac{1}{2}\mu C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \tag{13.1}$$

where $\mu$, the carrier mobility, decreases with an increase in temperature [33].

$$\mu = \mu_0\left(\frac{T}{T_0}\right)^{km} \tag{13.2}$$

where $km$ ranges from −1.2 to 2, and $T$ and $T_0$ are, respectively, the practical and reference temperatures.

$$V_T(T) = V_T(T_0) + \alpha(T - T_0) \tag{13.3}$$

where $\alpha$ is in the range of −0.5 mV/K to −3 mV/K.

In the case of digital cells, $V_{GS} >> V_T$, and thus mobility will dominate the thermal characteristics of the cell. Therefore, as temperature rises, the conductivity of all cells in the delay line will decline, which, in turn, will increase the delay of the cells. These events directly affect the resolution of the TDC. Stabilized performance requires eliminating the thermal dependence of the delay through the cells, as seen in Figure 13.4 [19].

The problem of the negative temperature sensitivity of conduction current is solved as depicted in Chen et al. [19]. The $V_T$ of the diode-connected transistors, P1 and N1, has an inverse relationship with temperature. As temperature increases, this drop in $V_T$ results in a rise in the terminal voltage of P3, thereby increasing its conduction current. The current-mirror configuration of P1, P2 and N1, N2 transfers this conduction current of P3 with a positive temperature coefficient to compensate for the loss in mobility in the inverter.

The value of the thermally independent $I_{D,P3}$ can be derived as

$$I_{D,P3} = \frac{1}{2}\mu_0 C_{ox}\left(\frac{W}{L}\right)(\alpha T_0)^2\left(1+\lambda V_{GS,P3}\right) \tag{13.4}$$

With this technique, thermally insensitive cells for delay-line techniques can be designed to give thermally insensitive TDC resolution.

### 13.4.2 Voltage Variations

Supply-voltage variations and process variations directly affect the delays of the elements used in the delay chain. The delay-locked loop (DLL) (Figure 13.5) performs time interpolation within the reference clock cycle. The DLL consists of a voltage-controlled delay line regulated by a control loop that forces it to span exactly one clock

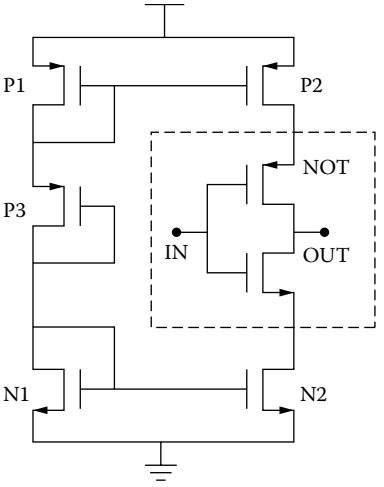



**FIGURE 13.4** Basic inverter delay cell with temperature compensation. (From Chen et al. [19]. © IEEE 2005. With permission.)

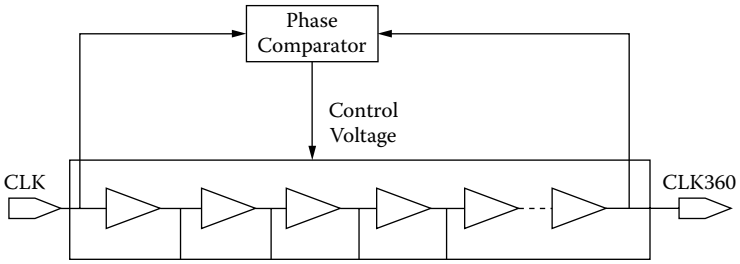



**FIGURE 13.5** Delay-locked loop. Total delay of the loop equals the clock period. Phase comparator circuit dynamically adjusts the control voltage to the delay cells to compensate for any supply or temperature variations.

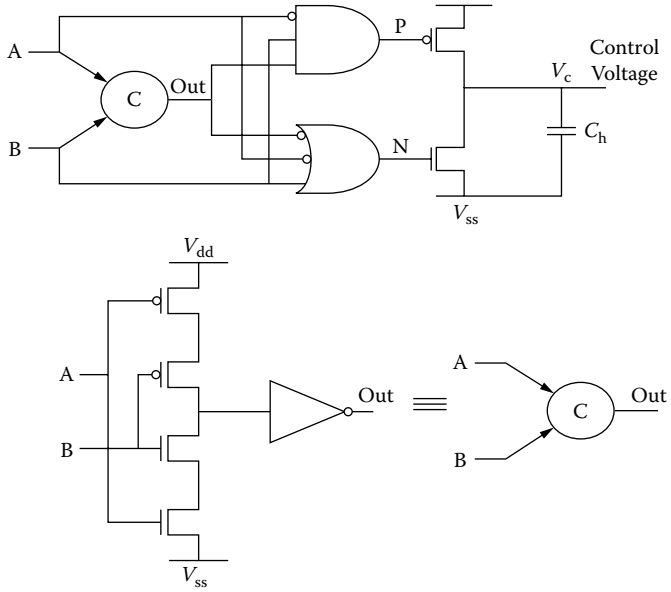



**FIGURE 13.6** Phase comparator and charge pump circuit. (From Santos et al. [34]. © IEEE 1996. With permission.)

cycle. The control loop comprises a phase detector (PD), a charge pump (CP), and a filter capacitor (Figure 13.6). The PD continuously measures the phase difference across the delay chain and directs the charge pump and filter circuit to adjust the control voltage accordingly. In this way, the delay of each delay cell is maintained at an average value equal to the clock period, divided by the number of delay cells in a chain.

The phase-detector circuit shown in Figure 13.6 distinguishes between the phase lag of one input with respect to the other. Unlike the phase-locked loop (PLL) circuit, two inputs to the PD circuit are correlated. The P and N pulses needed to drive the charge pump in Figure 13.6 are generated by a phase detector based on the Miller C-element [34].

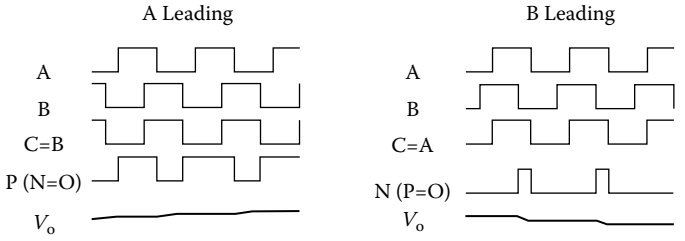



**FIGURE 13.7** Idealized waveforms of the phase detector circuit showing the proportional control. (From Santos et al. [34]. © IEEE 1996. With permission.)

Figure 13.7 shows the idealized waveforms for the complete phase-detector circuit. The output of the C-element follows whichever of the inputs, A or B, lags the other. Simple combinational logic elements generate P and N signals to control the pump that supplies the bias voltage for the delay chain. The circuit operates on the falling edge of the waveforms A and B. The proportional negative-feedback circuit ensures that, under lock, P stays at logic level 1 and N at 0.

## 13.5 NOISE ANALYSIS

A TDC-based quantizer introduces nonidealities due to noise and jitter, along with circuit nonuniformities. Noise is classified into deterministic noise and random noise. The source of deterministic noise is the periodic switching of the system clock; random noise, originating as a substrate and as supply noise, has a Gaussian probability distribution function (PDF). This section focuses on the latter.

The TDL's architecture relies on a stable, external crystal oscillator to provide a clock reference. Deviations in the placement of and routing delay for each delay element cause stationary INL and DNL variations [10]. They can be accounted for during the calibration phase. Furthermore, supply noise and substrate noise introduce time-variant oscillator jitter. The expression of phase jitter [35] is given by

$$\sigma_{\Delta\phi}^2 = \frac{\Gamma_{\text{rms}}^2 \overline{i_n}^2 / \Delta f}{2q_{\max}^2} \Delta T \tag{13.5}$$

where, $\sigma_{\Delta\phi}$ is the standard deviation in phase jitter due to a single white-noise source, $\Gamma_{\text{rms}}$ is the impulse-sensitivity function (ISF), $\overline{i_n}^2 / \Delta f$ is the single sideband power spectral-density of the noise-current source, $\Delta T$ is the measurement interval, and $q_{\max}$ is the maximum charge swing across the oscillator's output node.

The ISF for the ring oscillator is also given in Equation (13.6), from Hajimiri et al. [35],

$$\Gamma_{\text{rms}} = \sqrt{\frac{2\pi^2}{3\eta^2}} \frac{1}{N_{\text{r}}^{1.5}} \tag{13.6}$$

where $\eta$ is a technology-dependent proportionality constant and $N_{\text{r}}$ is the number of ring oscillator stages. ISF can be defined for any oscillator, and the corresponding phase noise calculated.

Phase jitter defines the uncertainty in the transition point for TDC output codes and also affects INL and DNL. For input signals with widths near the nominal transition levels, the TDC output switches randomly from one output code to an adjacent one [9]. The width of this uncertainty is driven by jitter, as denoted in Equation (13.6).

For the TDL architecture, this jitter directly translates into transition uncertainty. For a two-ring oscillator structure, uncertainty is cancelled out because the final TDC output code uses the difference between the two oscillators; it is equal to the noise present on the input signal.

The circuit introduces another level of uncertainty. The basic building blocks of all digital TDCs are delay cells and D-flip-flops. In asynchronous TDC circuits, the effect on the TDC's characteristics due to minimum setup and violations in hold time depends on the system's architecture, and the following section discusses the effects of the resulting meta-stability.

## 13.6 META-STABILITY IN DIGITAL LOGIC

TDC circuit building blocks contain D-flip-flops (DFF) and basic logic gates, such as inverter, Nand, and Xor. Digital sequential circuits involving DFF with asynchronous clock and data signals manifest metastable behavior when the metastability conditions [36] are satisfied. If the specifications for the DFF setup and hold time are not met, electronic noise dictates the DFF's output states, which entails erroneous readings of the counter. In the TDL architecture, this behavior causes bubbles in the thermometer code at the output of delay cells where the violations exist. Bubble correction logic easily corrects the output.

In two-ring oscillator topology, the circuit that compares the clocks may exhibit metastability occasioned by electronic noise. The output width is given in [10]. Two-ring-oscillator topology relies on precise phase crossover detection between two asynchronous ring-oscillator outputs, i.e., slow clock and fast clock. Phase crossover detection circuitry may exhibit metastable behavior. Due to the metastable nature of the DFF, clock crossover detection for the inputs with trailing edges or stop-signal alignment with the slow clock phase results in an uncertainty equal to the slow clock period. Two-ring-oscillator topology uses the following simple line equation to arrive at the input pulse width measurement $T_s$.[10].

Metastability may also exist at the crossover boundary between analog and digital circuits, where comparators convert analog voltage to a ToT (time over threshold) digital output, as discussed in detail in the next section. There is some minimum signal $V_L$ that digital circuits following a comparator can interpret unambiguously as either a 1 or 0. If the comparator gain is $A_0$, then inputs within $V_L/A_0$ of the threshold will produce indecisive outputs. For input signals uniformly distributed in voltage, the probability $p_i$ that there will be an indecisive comparator output in an encoder with quantization step $q$ is given in [37].

Achievement of a lower bit error rate (BER) is an optimization problem similar to that of an operational trans amplifier (OTA), needing higher gain and higher bandwidth. LVDS comparators at the input pads of the FPGA have adequate bandwidth (>100 MHz), high gain, and offer an excellent solution for the ToT application described in Section 13.7 [38].

## 13.7 TDC STATE OF THE ART

TDCs have been realized, with varying degrees of specifications, using CMOS, BiCMOS [39], ECL [40], rapid single flux quantum (RSFQ) electronics [16, 41, 42], SiGe, and FPGAs [3, 4]. Since RSFQ technologies operate only at cryogenic temperatures, they are not relevant for room-temperature detector applications. SiGe, BiCMOS, and ECL technologies are not very common, and sometimes it is hard to find a manufacturer offering these processes owing to relatively inexpensive and commercially available and widely used deep-submicron CMOS processes.

Deep-submicron CMOS technologies with high density, low power consumption, and low mass-production costs are very promising in realizing TDCs with fine resolutions, high dynamic ranges, and low power consumption compared to the conventional TAC approach. However, for research and experimental applications, the cost of designing and manufacturing a custom ASIC overwhelms the benefits offered by CMOS ASICs. This poses an even greater problem because the development process invariably involves at least two to three iterations of the design, based on prototype chip performances. The FPGAs offer an instant implementation and verification platform and extremely low prototype manufacturing costs.

Contemporary FPGAs designed in similar underlying CMOS processes come with most of the benefits associated with deep-submicron processes, and have the advantage of a considerably low design cost. A major limiting factor in FPGA-based TDCs is the lack of all the niceties of analog-based designs, which give control over voltage and temperature variations. This poses a challenge in implementing a TDC using an FPGA, because PVT variations limit the achievable linearity and resolution. The most common approach so far has been to achieve the best possible resolution by architectural modifications and by trial and error. The following sections focus on efforts to realize both of the popularly used ASIC- and FPGA-based TDCs.

### 13.7.1 FPGAs and CMOS-ASIC-Based TDCs

FPGAs mainly have been used in three types of applications. Firstly, their deployment focused on computationally intensive adaptive systems, such as real-time signal processing. The ability to configure the hardware so that it provides the most efficient path for the algorithm running at a given time can speed things up by orders of magnitude. Secondly, engineers turned to FPGAs in cases where the hardware needed updating in situ, particularly important in networking and aerospace, where systems can be in service for decades. Finally, ASIC designers could complete their work much more quickly using FPGAs as test beds, rather than waiting for test chips to be returned by the foundry [43].

FPGA-based TDC structures use a completely digital design incorporating a tapped delay line similar to that shown in Figure 13.8. Major design challenges faced during those implementations were the unpredictability of the FPGA-compiler's placing of the delay chain, the register structure of the fine TDC component, and PVT variations. Nowadays, dealing with PVT variations inside an FPGA is a greater challenge than the placements of the delay chain and the registers because of the improvements made in modern-day compilers [8]. An input clock with a known,

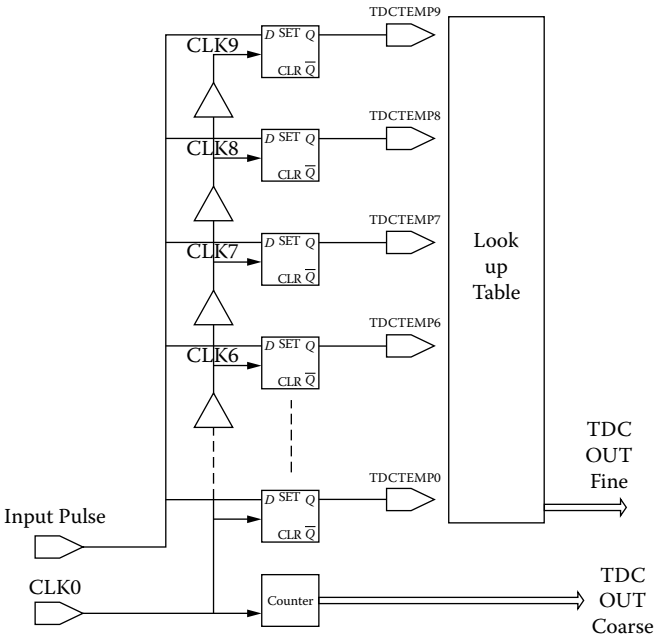



**FIGURE 13.8** TDC fine component implemented using the tapped-delay line. Clock is delayed using delay cells to obtain phases equally spaced in time. Latched DFF outputs provide a thermometer code that is converted to a binary number using a lookup table.

stable period provided an on-chip timing standard. The delay chain was made longer than one clock period so that the input was registered twice, first by cell number N1 and then by cell number N2, after exactly one clock period, *T*. The average bin size was equal to $T/(N2 - N1)$. Such an approach allows no control over the individual delay cells, although it provides a stable clock reference for compensation, thereby introducing an uncertainty equal to minimum of ±1 LSB. Synchronizing the coarse TDC counter with the fine TDC counter is very straightforward for this architecture, giving it a very large dynamic range.

QuickLogic FPGA with amorphous antifuse structure, pASIC architecture, has been employed to realize TDC [2]. For digitally coding measured time intervals, there are two tapped delay lines working in a differential mode, also known as VDL (Vernier delay line), described later. The first is created as a chain of the latch flip-flops, and the other as a chain of noninverting buffers. Hence, the basic delay cell contains one latch and one buffer. If the latch delay is longer than the buffer delay, the time quantization step (or the incremental resolution) of the TDC is determined by their difference, and can be made considerably shorter, by almost an order of magnitude, than the propagation time of the fastest gate in the FPGA logic block. An advantage of direct coding is its very short conversion time and very short dead time, which is equal to the readout time of the output data. The disadvantage of this structure is its very small (10 ns) dynamic range. Concatenating the coarse time stamp with the fine TDC value to increase the dynamic range is not straightforward because

the conversion time of a VDL is greater than the clock's period. Also, because of this difference a second event might enter the VDL even before it has finished converting the first one. Varner et al. [44] reported a TDC resolution of 2 ns using dual phases of a 250-MHz clock with an expected standard deviation of 0.6 ns.

These techniques have pushed the achievable resolution for an FPGA-based TDC to its architectural limits. VDL architecture is very promising, but its dynamic range and PVT variations on the order of the LSB limit its performance. Ideally, a TDC's resolution should be governed by the substrate and supply noise, and not its layout. Pushing the TDC resolution close to the ideal noise-limited value for an FPGA-based design is a challenge that is immediately applicable to PET and time-of-flight experiments. Certain architectures, such as the self-calibrating two-ring-oscillator- based technique described by Woodward et al. [37] are less susceptible to PVT variations, and come very close to a noise-limited performance.

### 13.7.2 Circuit Topologies

Several different circuit topologies have been used to implement a TDC. The resolution of a complete digital approach that uses delay cells and latches as primary circuit elements is limited by the clock's speed. The Vernier delay line (VDL) technique is the most promising of all, with an achievable resolution as fine as a few tens of picoseconds. Cyclic TDC has a comparable performance to the complete digital approach, but it has the disadvantage of larger dead times. The roadblocks to achieving noise-limited TDC performance are the PVT variations. The following section discusses the techniques established to realize fine TDC and to achieve better PVT performance.

#### 13.7.2.1 Tapped-Delay-Line Method

A complete digital-design approach usually encompasses coarse and fine components. The former is a binary counter running at the system clock speed that gives a time measurement accurate to the system clock's period. The latter is derived from several evenly spaced clock phases used to refine timing resolution. Latching the leading edge of the signal to be time-stamped to these clock phases allows us to extract timing information with a subclock-period accuracy. This technique of realizing a fine TDC component is often referred to as the tapped-delay-line method. Several design techniques have been incorporated to delay the clock with a predictable phase shift. One of them generates a delayed clock by having the clock buffered through logic gates or CMOS delay stages [45]. Another method uses RC delay lines to delay the signal with predictable phase. The achievable phase, equal to the propagation delay of the logic gate plus the routing delays, is a technology-dependent parameter and dictates the TDC resolution.

The tapped-delay-line (TDL) architecture derives several evenly spaced clock phases or delayed versions of the input pulse to refine timing resolution. Owing to its ease of implementation and reasonably higher resolution, many experiments have chosen this topology [1, 8–10, 37].

In the delayed-clock version, latching the leading edge of the signal to be time-stamped with these clock phases allows the extraction of timing information with subclock-period accuracy. This technique of realizing the fine TDC component often

is referred to as the TDL method. Similarly, an input pulse may be delayed using evenly spaced TDL, resulting in a thermometer-output code similar to the delayed-clock version. There's a very slight reduction in dynamic power dissipation with the delayed-input version as compared to the delayed-clock version, but for all practical purposes these two topologies are identical.

A complete digital-design approach [1] using TDL rests on basic FPGA building blocks. The major design challenges faced initially in its development were the unpredictability of the FPGA compiler placing the delay chain, the register structure of the fine-TDC component, and variations in process, voltage, and temperature (PVT). PVT variations pose a greater challenge than placing the delay chain and the registers. Control over placements of the logic cells is achievable due to improvements in modern-day compilers, which makes the implementation realizable [8–10]. An input clock with a known, stable period can be used to provide the on-chip timing standard. Such an approach lacks control over the individual delay cells, although it affords a stable clock reference for compensation, which introduces an uncertainty equal to a minimum of ±1 LSB (least-significant bit). Synchronizing the coarse TDC counter with the fine TDC counter is straightforward for this architecture, resulting in a dynamic range limited only by the number of bits in the counter.

Another concern in realizing FPGA-based TDC is the logic elements consumed for fine-TDC realization. TDL occupies more area on the FPGA, since utilizing the remaining logic elements to implement supporting logic in the TDL home logic array block (LAB) is not advisable, as it may result in higher DNL and INL due to routing congestion.

TDL resolution is equal to the minimum cell delay and is a parameter dependent on technology and FPGA speed (routing delays). The Altera Cyclone II family of FPGAs reported 1.2-ns resolution [8], and the Stratix II family has reported 625-ps resolution [9]. One of the earlier attempts at using Altera Acex 1K devices [1] showed an excellent resolution of 400 ps.

#### 13.7.2.2 Vernier Technique

The Vernier measurement technique, devised for measuring physical length, was extended in the CMOS domain for time measurement. For Vernier-based TDCs, time is resolved by two delay lines with very little difference in cell delay. The effective time-resolution equals the cell-delay difference [46]. Figure 13.9 shows its basic configuration. This delay line measures the time difference between rising edges of the inputs delayed and tapped by the two delay lines. The technique is based on a Vernier principle [47]. This delay line measures the difference in the timing of the rising edges of the inputs delayed and tapped by two delay lines [48]. The delay $t_1$ of a buffer in the upper delay chain is slightly greater than the delay $t_2$ of a buffer in the lower delay chain (Figure 13.9). As the CLK and INPUT signals propagate in their respective delay chains, the time difference between the CLK- and the INPUT-pulse declines in each Vernier stage. The position in the delay line at which the INPUT signal catches up with the CLK signal gives information about the measured time, with resolution $t_R$ equal to the difference between $t_1$ and $t_2$ [49].

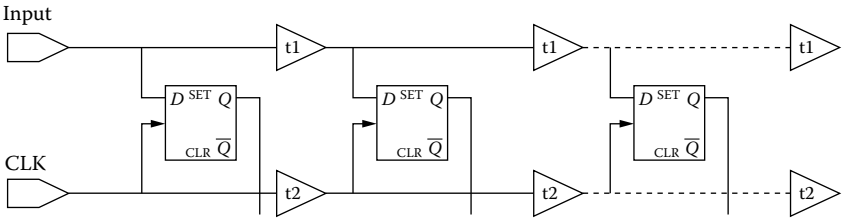



**FIGURE 13.9** Vernier delay line. Two TDLs are used to delay input and clock. Unit delays of the TDL differ very slightly by design, and the achievable resolution is the difference between those two delays. It Is equal to $t_1 - t_2$ for the shown VDL architecture.

In the literature [48], an Altera Acex 1K implementation of VDL showed LSB of 228 ps compared to 317-ps resolution achieved using Xilinx Virtex II devices. Xilinx devices, due to their PVT-controlled delay line available at the I/O pads, offer an attractive solution for TDL as well as VDL implementations.

Altera TDL and VDL implementations suffer from PVT effects. Slow varying PVT effects are acceptable for most coincidence spectroscopy experiments. A dynamic range of 10–100 ns is sufficient for most coincidence-spectroscopy applications. PVT variations do not affect the overall detector performance, because achievable coincidence resolution in those cases is limited by the detectors. VDL requires more than twice the logic elements as compared to TDL because it requires two TDLs.

For photodetectors such as photomultiplier tubes (PMT), silicon photomultipliers (SiPM), microchannel plates (MCP) along with LSO, LYSO, ZnO scintillators or modern CZT detectors, achievable coincidence timing resolution could range from 1 ps to 1000 ps [44]. TDC resolution desired for such detectors has to be at least Nyquist-limited or better. The TDC dynamic range required for these faster detectors is usually an order of magnitude higher than the TDC resolution. VDL- and TDL-architecture LSBs are found to be limited to 100 ps to 500 ps and are not suitable for such experiments. PVT variations of smaller than ±1 LSB over the dynamic could still be acceptable over the operating dynamic range.

#### 13.7.2.3 Cyclic TDC

Another delay-line structure is used in cyclic mode to achieve resolution better than the system's clock speed. The delay line is composed of an even number of inverter gates. As the input pulse circulates in this cyclic delay line, its width shrinks by a specific amount per cycle, equal to the propagation delay of the inverter, $t_p$. A counter is incremented at the end of each cycle with a final value of $N_c$ until the pulse diminishes completely. Figure 13.10 shows such a structure. Pulse width is simply given by Equation (13.7),

$$T_w = N_c \times t_p \tag{13.7}$$

Figure 13.10 is a block diagram of the two-ring-oscillator technique. Oscillator frequencies intentionally are designed to differ very slightly and rely on dissimilarities caused by different load capacitances on the second stage of the oscillators, and

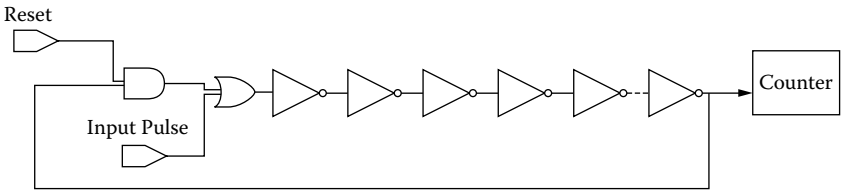



**FIGURE 13.10** Cyclic TDC structure. Every excursion of an input pulse around the inverter chain shrinks the pulse by an amount equal to the propagation delay ($t_p$) through the inverter. Counter at the end counts the total number of excursions ($N_c$) until the pulse disappears completely, giving the pulse width equal to $N_c \times t_p$.

disparities in placement and routing delays. The achievable resolution is equal to this difference, which varies between FPGA families. Multiple TDC channels can be reproduced by ensuring that the load capacitance dominates the differences in placement and routing. Due to its differential architecture, this scheme is less susceptible to PVT variations. Achievable resolution and TDC linearity is very close to the noise-limited performance. This design is also attractive because it is self-calibrating and addresses most of the issues associated with TDC performance. The usage of this architecture is limited by the time-dependent jitter of the ring oscillators, which tends to restrict the system's usable dynamic range. But, with careful design and a stable external oscillator, the dynamic range can be extended, and then is limited only by the number of bits in a counter.

A time-to-amplitude converter, in conjunction with a conventional ADC, is an analog way for obtaining precision time measurements. However, this technique is gradually being replaced by the all-digital methods discussed previously, owing to common problems associated with analog solutions, such as noise, cross talk, and integral and differential nonlinearity.

In ToF applications where larger dynamic range is desired (usually in the range of 0.1–10 μs), PVT variations affect the TDC performance. On-chip timing standards such as crystal oscillators with accuracy better than or equal to 1 ppm are used to partially circumvent those issues. Periodic calibration regimens could be employed to arrive at LSB and linearity calibration numbers right before or during the experiment. Calibration cycles could be made part of the data-acquisition routines, but may add to the system dead time, since most of the calibration cycles require external equipment such as a pulse generator, data-acquisition-and-control system, etc. The following section reviews an architecture that is relatively immune to PVT, promises sub-100-ps resolution, and has virtually unlimited dynamic range.

#### 13.7.2.4 Two-Ring Oscillator-Based Technique

The two-ring oscillator technique measures either the interval between a start and a stop pulse or the width of an input pulse. The TDC's output code is derived in two steps: a coarse time stamp, with a resolution equal to the period of the slow oscillator clock, and a fine time stamp using the two-ring oscillator scheme described here.

Oscillator frequencies intentionally are designed to differ very slightly; they rely on dissimilarities caused by different load capacitances on the second stage of the

oscillators, and on disparities in their placement and routing delays [10]. The assignment-based approach used in a tapped-delay-line architecture dictates the placing of the ring oscillators' logic cells. Timing analysis and simulation predict differences in the ring-oscillators' frequencies; the achievable resolution equals this difference. Multiple TDC channels are reproduced by ensuring that load capacitance dominates the differences in their placement and routing. This technique requires fewer than 10 logic elements to realize two-ring oscillators (Figure 13.11). The logic elements near the oscillators are utilized for realizing the Gray counters, ensuring optimal usage of the real estate. Ring-oscillator jitter increases proportionally with the measurement time [35]. This limits the dynamic range of the fine TDC component to a few times the slowest ring-oscillator period. This problem is easily circumvented by extending the dynamic range using an external oscillator with jitter within a few picoseconds range. The present state of the art pushes the achievable resolution very close to noise-limited numbers. Future detectors with timing resolutions of a few picoseconds will require better than present state-of-the-art electronics. FPGA-based TDCs may not be the right choice for those high-speed detectors.

An analog method for precisely measuring time is by combining time-to-amplitude converters with a conventional ADC. The digital techniques outlined previously are gradually replacing this technique because of problems common in analog solutions, such as noise, cross talk, and integral and differential nonlinearity.

The main contributors to the FoM are the resolution and dynamic power dissipation. The FoM achieved in all the topologies is more than adequate for the experiments, as the power dissipation in the fine TDC components is only a few percent of the total signal-processing system power budget. TDL and VDL fine TDC architectures are continuously running at system clock speed with continuous dynamic power dissipation. Two-ring oscillators are turned on only during the measurement and calibration phase and are off otherwise and have very low dynamic power dissipation.

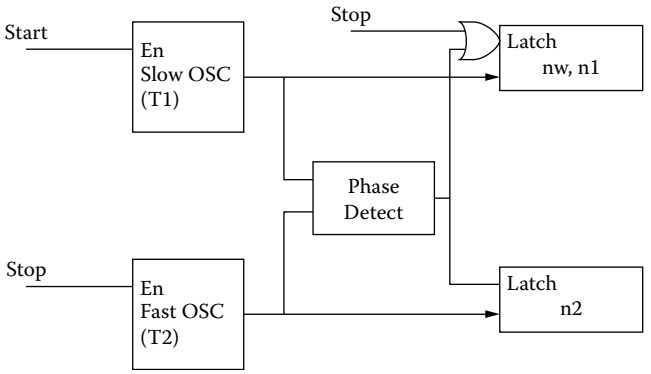



**FIGURE 13.11** Two-ring-oscillator topology. Two-ring oscillators are designed with slightly different time periods $T_1$ and $T_2$. Resolution achieved is $T_1 - T_2$. Time interval measured is the time between start and stop pulses. Each pulse starts its respective oscillators. The fast oscillator catches up with the slow oscillator, and phase-detect circuitry detects this crossover. (From Junnarkar et al. [10]. © IEEE 1998. With permission.)

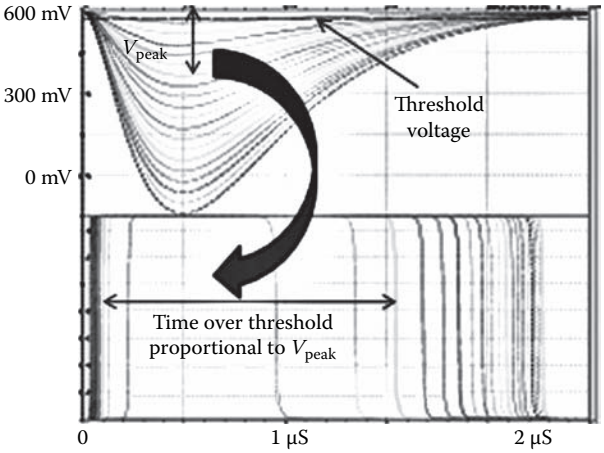



**FIGURE 13.12** Time-over-threshold simulation. Amplitude sweep shown as v(inputp), comparator threshold set using a DAC at vinput(n) and the corresponding PWM output v(recdoutbar_out).

Architectural techniques have pushed the achievable resolutions to noise-dominated limits, while the FoM has seen almost an exponential growth over the past decade.

ADC realization using the ToT technique is a promising architecture for radiation experiments. Unipolar shaper output is applied to a positive input of a comparator. A DC voltage threshold is applied to the negative input using a commercial digital-to-analog converter (DAC). ToT is given by the width of the comparator digital output pulse, also called a pulse-width-modulated (PWM) output. ToT has a logarithmic relationship with the analog signal peak amplitude. Simulation results with an Altera LVDS comparator SPICE model was used to evaluate the feasibility of the ToT technique using FPGA. A linear sweep was applied to the unipolar Gaussian-shaped analog signal amplitude (vpeak), with peaking time of 1 μs, as shown in Figure 13.12. The DAC threshold is held constant during the amplitude sweep, and the corresponding PWM output, shown in Figure 13.12 shows input charge for each step in the sweep versus the PWM output width. This solution serves as a very-high-resolution ADC, but is limited to larger peaking times.

## 13.8 CONCLUSION

This chapter presented TDC circuit implementations using innovative architectures. Their applications included positron emission tomography (PET) scanners and time-of-flight (TOF) experiments discussed in depth. TDC design specifications and the performance of TDL architectures in those systems were presented along with overall system performance.

Part of the research work presented in this chapter included a literature review of the TDC state of the art, and this was presented in the first few sections. Conventional methods of TDC realization included the use of CFD, TAC, and ADC to extract tim-

ing information. As the technology progressed, this was followed by many CMOS-ASIC- and FPGA-based architectures.

CMOS ASIC solutions for TDC circuits are desirable for many experiments because of their analog circuit design capabilities. For example, a charge-pump design implemented for a voltage-controlled delay line provides excellent compensation technique against process, voltage, and temperature variations [34]. Limitations of these CMOS ASIC solutions included the long timeline from concept to final product, as well as the high costs associated with custom CMOS fabrication runs. These circuits were also limited in their use to the specific applications they were designed for.

There were few attempts at reconfigurable FPGA-based implementations of TDC circuits prior to the work presented in this chapter. They show a natural evolution, which came with shrinking technology feature size. Most of the research on FPGA-based TDC faced challenges posed by limited control offered by the design tools and by architectural roadblocks, some of which were not well understood at that time. The work presented here studied those limitations in detail. A method was devised for TDC characterization that helped in understanding those issues.

This chapter presented research work undertaken in the pursuit of better performance while keeping the PET and time-of-flight experimental requirements in mind. The PET experiment titled Rat conscious animal positron emission tomography (RatCAP) uses 384 LSO-APD detectors coupled one-to-one ASIC channels. ASIC granularity is 32 channels and employs a 32 to 1 priority encoding scheme to encode geographical locations on a $4 \times 8$ LSO-APD array. Such 12-detector ASIC channels are served using single TSPM for awake-animal experiments. The total power dissipation of the TSPM, including 12 TDCs, signal-processing logic, and data acquisition with three G-Link channels (one channel in and two out), is close to 10 W. Less than 1% of the power is dissipated for TDC and signal processing, while the remaining 99% is consumed for the G-Links. RatCAP technology inspired many other PET detectors, of which four were for simultaneous PET-MRI applications and one for wrist detector. The compactness of the RatCAP technology, along with efforts undertaken to make the technology MRI compatible, yielded PET detectors that are compatible with four different MRI scanners. There are three applications for small-animal PET scans, and one experiment is designed for breast cancer detection in situ with an Aurora Inc. MRI scanner. TDC designed using FPGA technology is the workhorse of the signal processing and data acquisition for those PET detectors ranging from small-animal to human applications.

TDL architecture is currently being used in the experiments and meets the design specifications for small-animal scanners. As the detector count increases for future applications, the TDL architecture may be adequate, but ring-oscillator-based architectures provide a more compact design. The self-calibrating nature of the ring oscillator technique is more desirable for multiple reasons. This architecture requires minimal area compared to other presented architectures. It has a FoM that is far superior to its contemporary architectures. This has a greater impact on overall system power dissipation. The two-ring-oscillator technique has a slight disadvantage of higher bit count per event due to higher resolution. The resolution of 625 ps achieved using TDL architecture is adequate for the RatCAP-technology-based detectors, since the coincidence-timing resolution is limited by the detector used.

The future of small-form-factor PET scanners is the ToF genre of detectors, given that SiPM detectors promise timing in the picosecond range. Two-ring-oscillator-based architecture is ready to use for such futuristic scanners.

This work was published in three journal articles that also contributed significantly to the work presented in this chapter. The first publication provided TDC realized using an Altera Stratix II FPGA with 625-ps resolution. It showed excellent DNL and INL performance. The second article showed that most of the issues revealed during the first TDC implementation were well understood and resolved, including calibration, PVT variations, nonlinearity, and better resolution. The performance of self-calibrating two-ring-oscillator-based topology showed a resolution of 41 ps, rivaling or in some cases bettering CMOS ASIC counterparts. The third paper presented a summary of the research work. During this venture, it was observed that there was no single document that encompasses TDC circuits. The third journal article provided a comprehensive review of TDC implementations in the literature and also provided a theoretical basis for analyzing the performance of the TDC circuits in general.

TDL architecture suffered from noise from the crystal oscillator's phase, the supply, and the substrate. It is possible to design a TDL with a very low jitter (few ps) and very stable oscillator with close to 1 ppm accuracy. Therefore, to ensure TDL linearity, it is important to select oscillators with low jitter and low long-term drift. In the signal-delayed version of the TDL, supply- and substrate-noise impair the signal's edges, where the oscillator-phase noise introduced stationary uncertainty per measurement cycle. PVT variations in Altera FPGAs limited their usage to coincidence spectroscopy applications, where long-term relative stability is not a big restriction. ToF experiments require long-term stability, and for those applications Xilinx FPGAs with PVT-compensated delay lines offer more robust TDL and VDL architectures [48].

Also discussed in this chapter were the basic TDC architectures, giving analytical expressions for corresponding characteristics. The predominant TDC architectures in the literature are comprised of TDL, Cyclic, VDL, and self-calibrating two-ring oscillator architectures. Their performance was compared by normalizing their dynamic range. The corresponding FoMs were calculated in an attempt to monitor historical trends over the past decade. At a first glance, TDC performance has seen an exponential growth since its early antifuse structure implementation [2] in an FPGA. Part of the performance improvement is credited to the natural technology evolution, and part of the credit lies with the design techniques. Self-calibrating two-ring oscillator topology has superior FoM compared to contemporary implementations. Also, this structure outlines the PVT performance along with excellent DNL (<0.5 LSB) and INL (<1 LSB) with very linear behavior.

In conclusion, TDCs are crucial components of the front-end electronics signal path for radiation detectors. This chapter has presented innovative architectures for realizing TDC circuits using readily available technology platform, i.e., FPGAs. The performance reported in the chapter parallels its CMOS ASIC counterparts. FPGA-based or even CMOS-ASIC-based circuits find a noise-limited resolution barrier that tends to lie in the range of a few tens of picoseconds. There is a need for research and development to realize finer-resolution TDC circuits and low-phase noise oscillator circuits to pursue applications in the 1-ps timing regime.

## ACKNOWLEDGMENTS

The author wishes to thank Dr. Paul O'Connor, from Brookhaven National Laboratory (BNL), who has been a source of inspiration for the work on TDCs. Discussions with Dr. O'Connor and his suggestions not only added value to the chapter, contributing directly to the work, but also helped formulate the theory for TDC circuits. The field of FPGA-based TDCs is relatively uncharted territory, and to sail through it and achieve the results presented here would not have been possible without the constant encouragement and guidance of Dr. Rejean Fontaine from the Université De Sherbrooke. Also, this work was made possible because of the support of funding from the Department of Energy, Brookhaven National Laboratory's Instrumentation Division, and the RatCAP group. RatCAPpers have generously supported the TDC research and development work. Dr. Veljko Radeka, Dr. Craig Woody, Dr. David Schlyer, and Dr. Paul Vaska at BNL have encouraged the realization of a picoseconds-resolution TDC designed using FPGA.

## REFERENCES

1. Wu, J., et al. 2003. Firmware-only implementation of time-to-digital converter (TDC) in field-programmable gate array (FPGA). *IEEE NSS 2003 Conference Record* 1:177–181.
2. Kalisz, J., et al. 1997. Field-programmable gate-array-based time-to-digital converter with 200-ps resolution. *IEEE Trans. Instrumental Meas*. 46:51–55.
3. Bogdan, M., et al. 2005. A 96-channel FPGA-based time-to-digital converter. *Nuclear Instrumental Methods A* 554:444–457.
4. Karadamoglou, K., et al. 2004. An 11-bit high-resolution and adjustable-range CMOS time-to-digital converter for space science instruments. *IEEE J. Solid-State Circuits* 39:214–222.
5. Kleinfelder, S., et al. 1991. MTD 132: A new sub-nanosecond multi-hit CMOS time-to-digital converter. *IEEE Trans. Nuclear Sci*. 38:97–101.
6. Radeka, V. 1988. Low-noise techniques in detectors. *Ann. Rev. Nuclear Part. Sci*. 38:217–277.
7. Leroux, J.-D., et al. 2004. Time discrimination techniques using artificial neural networks for positron emission tomography. *IEEE Nuclear Science Symposium Conference Record* 4:2301–2305.
8. Junnarkar, S. S., et al. 2005. An FPGA-based 12-channel TDC and digital signal processing module for the RatCAP scanner. *IEEE Nuclear Science Symposium Conference Record* 2:919–923.
9. Junnarkar, S. S., et al. 2008. Next generation of real time data acquisition, calibration and control system for the RatCAP scanner. *IEEE Trans. Nuclear Science* 55:220–224.
10. Junnarkar, S. S., et al. 2009. FPGA based self calibrating 40 picosecond resolution, wide range time to digital converter. *IEEE Trans. Nuclear Science* 56:2374–2379.
11. Mota, M., et al. 2001. A flexible multi-channel high-resolution time-to-digital converter ASIC. *IEEE Nuclear Science Symposium Conference Record* 2:9/155–9/159.
12. Mota, M., et al. 1998. A four channel self-calibrating high resolution time to digital converter. *IEEE International Conference on Electronics, Circuits and Systems* 1:409–412.
13. Chen, C., et al. 2003. Front end electronics for the CDF-II time-of-flight system. *IEEE Trans. Nuclear Science* 50:2486–2490.
14. Mota, M., et al. 1999. A high-resolution time interpolator based on a delay locked loop and an RC delay line. *IEEE J. Solid-State Circuits* 34:1360–1366.

15. Christiansen, J., et al. 2000. A data driven high performance time to digital converter. Paper presented at 6th Workshop on Electronics for LHC Experiments, Krakow, Poland, 169–173.
16. Swann, B. K., et al. 2004. A 100-ps time-resolution CMOS time-to-digital converter for positron emission tomography imaging applications. *IEEE J. Solid-State Circuits* 39:1839–1852.
17. Moses, W. W., et al. 1999. Prospects for time-of-flight PET using LSO scintillator. *IEEE Trans. Nuclear Science* 46:474–478.
18. Staszewski, R. B., et al. 2004. All-digital TX frequency synthesizer and discrete-time receiver for Bluetooth radio in 130-nm CMOS. *IEEE J. Solid-State Circuits* 39:2278–2291.
19. Chen, C.-C., et al. 2005. A precise cyclic CMOS time-to-digital converter with low thermal sensitivity. *IEEE Trans. Nuclear Science* 52:834–838.
20. Ramakrishnan, V., et al. 2006. A wide-range high-resolution compact CMOS time to digital converter. In *Proceedings of the 19th International Conference on VLSI Design*. Washington, DC: IEEE Computer Society, 197–202.
21. Lee, L.-M., et al. 2006. A sub-10-ps multiphase sampling system using redundancy. *IEEE J. Solid-State Circuits* 41:265–273.
22. Staszewski, R. B., et al. 2006. 1.3 V 20 ps time-to-digital converter for frequency synthesis in 90-nm CMOS. *IEEE Trans. Circuits and Systems II: Express Briefs* 53:220–224.
23. Henry, C., et al. 2003. CMOS built-in test architecture for high-speed jitter measurement. Paper presented at International Test Conference (ITC 2003), Charlotte, NC.
24. Vaska, P., et al. 2004. RatCAP: Miniaturized head-mounted PET for conscious rodent brain imaging. *IEEE Trans. Nuclear Science* 51:2718–2722.
25. Pratte, J.-F., et al. 2004. Design and performance of 0.18 μm CMOS charge preamplifier for APD-based PET scanners. *IEEE Trans. Nuclear Science* 51:1979–1985.
26. Chatziioannou, A. F., et al. 2002. Molecular imaging of small animals with dedicated PET tomographs. *European Journal of Nuclear Medicine* 29:98–114.
27. Momosaki, S., et al. 2004. Rat-PET study without anesthesia: Anesthetics modify the dopamine D1 receptor binding in rat brain. *Synapse* 54:207–213.
28. Tsukada, H., et al. 2001. Ketamine alters the availability of striatal dopamine transporter as measured by [C-11]beta-CFT and [C-11]beta-CIT-FE in the monkey brain. *Synapse* 42:273–280.
29. Mitra, S., et al. 2006. Depth profiling of C and other light elements, in situ, using the associated particle technique. Proposal questionnaire. BNL, Laboratory Directed Research and Development Program, Brookhaven National Laboratory, Upton, NY.
30. Jones, K. M., et al. 1975. An associated particle system for evaluation of 15 MeV neutron beam shield and collimator designs. *Phys. Med. Biol*. 20:131–135.
31. Beyerle, A., et al. 1990. Design of an associated particle imaging system. *Nuclear Instrumental Methods A* 299:458–462.
32. Haysladen, P. A., et al. 2005. An alpha particle detector for a portable neutron generator for the Nuclear Materials Identification System (NMIS). *Nuclear Instrumental Methods B* 241:835–838.
33. Filanovsky, I. M., et al. 2001. Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits. *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications* 48:876–884.
34. Santos, D. M., et al. 1996. A CMOS delay locked loop and sub-nanosecond time-to-digital converter chip. *IEEE Trans. Nuclear Science* 43:1717–1719.
35. Hajimiri, A., et al. 1999. Jitter and phase noise in ring oscillators. *IEEE J. Solid-State Circuits* 34:790–804.
36. "A Metastability Primer," Application Note: AN219, Phillips Semiconductors, November 15, 1989.

37. Woodward, C., et al. 1975. A monolithic voltage-comparator array for A/D converters. *IEEE J. Solid-State Circuits* 10:392–399.
38. *IEEE Solid State Circuits Magazine*.
39. McDonald, J. F., et al. Reconfigurable FPGAs in the 1-20 GHz band with HBT BiCMOS. Rensselaer Polytechnic Institute, Troy, N.Y. http://www.rpi.edu.
40. Räisänen-Ruotsalainen, E., et al. A high resolution time-to-digital converter based on time-to-voltage interpolation. Electronics Laboratory, Department of Electrical Engineering, Oulu University, Oulu, Finland.
41. Kaplan, S. B., et al. 2001. A prescaler circuit for a superconductive time-to-digital converter. *IEEE Trans. Applied Superconductivity* 11:513–516.
42. Mukhanov, O. A., et al. 1997. Time-to-digital converters based on RSFQ digital counters. *IEEE Trans. Applied Superconductivity* 7:2669–2672.
43. Baines, S., et al. 2005. Never say die. *IEE Review* 51 (no. 6):44–48.
44. Varner, G. S., et al. 2006. The modern FPGA as discriminator, TDC and ADC. IOP Publishing Ltd. and SISSA. http://www.iop.org/EJ/jinst/.
45. Rahkonen, T. E., et al. 1993. The use of stabilized CMOS delay lines for the digitization of short time intervals. *IEEE J. Solid-State Circuits* 28:887–894.
46. Chen, P., et al. 2005. A monolithic Vernier-based time-to-digital converter with dual PLLs for self-calibration. In *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference*, 321–324.
47. Vernier Scale, http://en.wikipedia.org/wiki/Vernier_scale.
48. Amiri, A. M., et al. 2009. A multihit time-to-digital converter architecture on FPGA. *IEEE Trans. Instrumentation and Measurement* 58:530–540.
49. Dudek, P., et al. 2000. A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line. *IEEE J. Solid-State Circuits* 35:240–247.
50. Altera Corp. Stratix II device family data sheet. www.altera.com.
51. Beyerle, A., et al. 1990. Design of an associated particle imaging system. *Nuclear Instrumental Methods A* 299:458.
52. Bogdan, M., et al. 2004. A new time-to-digital converter for the central tracker of the colliding detector at Fermilab. *IEEE 2004 NSS Conference Record* 3:1666–1668.
53. Chen, C.-C., et al. 2005. A time-to-digital-converter-based CMOS smart temperature sensor. *IEEE International Symposium on Circuits and Systems*. 1:560–563.
54. Haysladen, P. A., et al. 2005. An alpha particle detector for a portable neutron generator for the Nuclear Materials Identification System (NMIS). *Nuclear Instrumental Methods B* 241:835–838.
55. Dudek, P., et al. 2000. A high-resolution CMOS time-to-digital converter utilizing a Vernier delay line. *IEEE Trans. Solid-State Circuits* 35:240–247.
56. Lin, J., et al. 2004. A PVT tolerant 0.18MHz to 600MHz self-calibrated digital PLL in 90nm CMOS process. *ISSCC Solid-State Circuits Conference 2004, Digest of Technical Papers*1:488–541.
57. Lin, Y., et al. 2005. Circuits and architectures for field programmable gate array with configurable supply voltage. *IEEE Trans. VLSI Systems* 13:1035–1047.
58. Mukhanov, O. A., et al. 1997. Time-to-digital converters based on RSFQ digital counters. *IEEE Trans. Applied Superconductivity* 7:2669–2672.
59. van de Plassche, R. 2003. *CMOS integrated analog-to-digital and digital-to-analog converters*. Dordrecht: Kluwer Academic Press.
60. Pratte, J.-F., et al. 2004. Design and performance of 0.18-μm CMOS charge preamplifier for APD-based PET scanners. *IEEE Trans. Nuclear Sci*. 51:1979–1985.
61. Leroux, J.-D., et al. 2003. Time determination of BGO-APD detectors by digital signal processing for positron emission tomography. *Nuclear Science Symposium Conference Record* 3:1723–1727.

# Index

## D



## E



## F



## G

## R



## S

## T

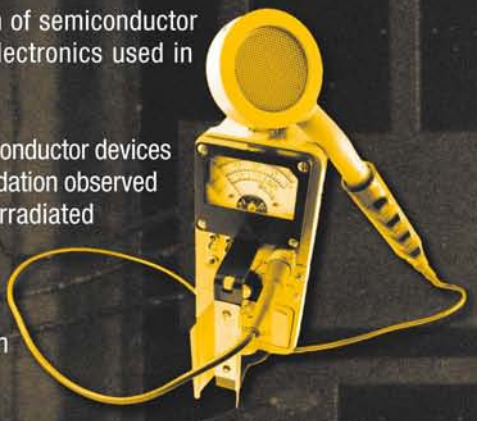
Electronics

# Electronics for Radiation Detection

There is a growing need to understand and combat potential radiation damage problems in semiconductor devices and circuits. Assessing the billion-dollar market for detection equipment in the context of medical imaging using ionizing radiation, **Electronics for Radiation Detection** presents valuable information that will help integrated circuit (IC) designers and other electronics professionals take full advantage of the tremendous developments and opportunities associated with this burgeoning field.

*Assembling contributions from industrial and academic experts, this book—*

- Addresses the state of the art in the design of semiconductor detectors, integrated circuits, and other electronics used in radiation detection
- Analyzes the main effects of radiation in semiconductor devices and circuits, paying special attention to degradation observed in MOS devices and circuits when they are irradiated
- Explains how circuits are built to deal with radiation, focusing on practical information about how they are being used, rather than mathematical details



Radiation detection is critical in space applications, nuclear physics, semiconductor processing, and medical imaging, as well as security, drug development, and modern silicon processing techniques. The authors discuss new opportunities in these fields and address emerging detector technologies, circuit design techniques, new materials, and innovative system approaches.

Aimed at postgraduate researchers and practicing engineers, this book is a must for those serious about improving their understanding of electronics used in radiation detection. The information presented here can help you make optimal use of electronic detection equipment and stimulate further interest in its development, use, and benefits.

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
270 Madison Avenue
New York, NY 10016
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K10847
ISBN: 978-1-4398-1648-6
90000
9 781439 816486