

**THE
ESSENTIAL
PHYSICS OF
MEDICAL
IMAGING
Second Edition**



**JERROLD T. BUSHBERG
J. ANTHONY SEIBERT
EDWIN M. LEIDHOLDT, JR.
JOHN M. BOONE**



THE ESSENTIAL PHYSICS OF MEDICAL IMAGING

SECOND EDITION

JERROLD T. BUSHBERG, PHD

*Clinical Professor of Radiology
University of California, Davis
Sacramento, California*

J. ANTHONY SEIBERT, PHD

*Professor of Radiology
University of California, Davis
Sacramento, California*

EDWIN M. LEIDHOLDT, JR., PHD

*Associate Clinical Professor of Radiology
University of California, Davis
Sacramento, California*

JOHN M. BOONE, PHD

*Professor of Radiology
University of California, Davis
Sacramento, California*



LIPPINCOTT WILLIAMS & WILKINS

A Wolters Kluwer Company

Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

Acquisitions Editor: Joyce-Rachel John
Developmental Editor: Anne Snyder
Production Editor: Tony DeGeorge
Manufacturing Manager: Tim Reynolds
Cover Designer: John M. Boone and Jerrold T. Bushberg
Compositor: Lippincott Williams & Wilkins Desktop Division
Printer: Edwards Brothers

© 2002 by LIPPINCOTT WILLIAMS & WILKINS
530 Walnut Street
Philadelphia, Pa 19106 USA
LWW.com

All rights reserved. This book is protected by copyright. No part of this book may be reproduced in any form or by any means, including photocopying, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotations embodied in critical articles and reviews. Materials appearing in this book prepared by individuals as part of their official duties as U.S. government employees are not covered by the above-mentioned copyright.

Printed in the USA

Library of Congress Cataloging-in-Publication Data

The essential physics of medical imaging / Jerrold T. Bushberg [et al.].—2nd ed.

p. ; cm

Includes bibliographical references and index.

ISBN-13: 978-0-683-30118-2

ISBN-10: 0-683-30118-7

1. Diagnostic imaging. 2. Medical physics. I. Bushberg, Jerrold T.

[DNLM: 1. Diagnostic Imaging-methods. WN 200 E78 2001]

RC78.7.D53 E87 2001

616.07'54—dc21

2001041711

Care has been taken to confirm the accuracy of the information presented and to describe generally accepted practices. However, the authors, editors, and publisher are not responsible for errors or omissions or for any consequences from application of the information in this book and make no warranty, expressed or implied, with respect to the currency, completeness, or accuracy of the contents of the publication. Application of this information in a particular situation remains the professional responsibility of the practitioner.

The authors, editors, and publisher have exerted every effort to ensure that drug selection and dosage set forth in this text are in accordance with current recommendations and practice at the time of publication. However, in view of ongoing research, changes in government regulations, and the constant flow of information relating to drug therapy and drug reactions, the reader is urged to check the package insert for each drug for any change in indications and dosage and for added warnings and precautions. This is particularly important when the recommended agent is a new or infrequently employed drug.

Some drugs and medical devices presented in this publication have Food and Drug Administration (FDA) clearance for limited use in restricted research settings. It is the responsibility of the health care provider to ascertain the FDA status of each drug or device planned for use in their clinical practice.

10 9

CONTENTS

Preface xv
Acknowledgments xvii
Foreword xix

SECTION I: BASIC CONCEPTS 1

Chapter 1: Introduction to Medical Imaging 3

- 1.1 The Modalities 4
- 1.2 Image Properties 13

Chapter 2: Radiation and the Atom 17

- 2.1 Radiation 17
- 2.2 Structure of the Atom 21

Chapter 3: Interaction of Radiation with Matter 31

- 3.1 Particle Interactions 31
- 3.2 X- and Gamma Ray Interactions 37
- 3.3 Attenuation of X- and Gamma Rays 45
- 3.4 Absorption of Energy from X- and Gamma Rays 52
- 3.5 Imparted Energy, Equivalent Dose, and Effective Dose 56

Chapter 4: Computers in Medical Imaging 61

- 4.1 Storage and Transfer of Data in Computers 61
- 4.2 Analog Data and Conversion between Analog and Digital Forms 66
- 4.3 Components and Operation of Computers 70
- 4.4 Performance of Computer Systems 78
- 4.5 Computer Software 79
- 4.6 Storage, Processing, and Display of Digital Images 82

SECTION II: DIAGNOSTIC RADIOLOGY 95

Chapter 5: X-ray Production, X-ray Tubes, and Generators 97

- 5.1 Production of X-rays 97

viii *Contents*

- 5.2 X-ray Tubes 102
- 5.3 X-ray Tube Insert, Tube Housing, Filtration, and Collimation 113
- 5.4 X-ray Generator Function and Components 116
- 5.5 X-ray Generator Circuit Designs 124
- 5.6 Timing the X-ray Exposure in Radiography 132
- 5.7 Factors Affecting X-ray Emission 135
- 5.8 Power Ratings and Heat Loading 137
- 5.9 X-ray Exposure Rating Charts 140

Chapter 6: Screen-Film Radiography 145

- 6.1 Projection Radiography 145
- 6.2 Basic Geometric Principles 146
- 6.3 The Screen-Film Cassette 148
- 6.4 Characteristics of Screens 149
- 6.5 Characteristics of Film 157
- 6.6 The Screen-Film System 163
- 6.7 Contrast and Dose in Radiography 164
- 6.8 Scattered Radiation in Projection Radiography 166

Chapter 7: Film Processing 175

- 7.1 Film Exposure 175
- 7.2 The Film Processor 178
- 7.3 Processor Artifacts 181
- 7.4 Other Considerations 183
- 7.5 Laser Cameras 184
- 7.6 Dry Processing 184
- 7.7 Processor Quality Assurance 186

Chapter 8: Mammography 191

- 8.1 X-ray Tube Design 194
- 8.2 X-ray Generator and Phototimer System 204
- 8.3 Compression, Scattered Radiation, and Magnification 207
- 8.4 Screen-Film Cassettes and Film Processing 212
- 8.5 Ancillary Procedures 219
- 8.6 Radiation Dosimetry 222
- 8.7 Regulatory Requirements 224

Chapter 9: Fluoroscopy 231

- 9.1 Functionality 231
- 9.2 Fluoroscopic Imaging Chain Components 232
- 9.3 Peripheral Equipment 242
- 9.4 Fluoroscopy Modes of Operation 244
- 9.5 Automatic Brightness Control (ABC) 246
- 9.6 Image Quality 248
- 9.7 Fluoroscopy Suites 249
- 9.8 Radiation Dose 251

Chapter 10: Image Quality 255

- 10.1 Contrast 255
- 10.2 Spatial Resolution 263
- 10.3 Noise 273
- 10.4 Detective Quantum Efficiency (DQE) 283
- 10.5 Sampling and Aliasing in Digital Images 283
- 10.6 Contrast-Detail Curves 287
- 10.7 Receiver Operating Characteristics Curves 288

Chapter 11: Digital Radiography 293

- 11.1 Computed Radiography 293
- 11.2 Charged-Coupled Devices (CCDs) 297
- 11.3 Flat Panel Detectors 300
- 11.4 Digital Mammography 304
- 11.5 Digital versus Analog Processes 307
- 11.6 Implementation 307
- 11.7 Patient Dose Considerations 308
- 11.8 Hard Copy versus Soft Copy Display 308
- 11.9 Digital Image Processing 309
- 11.10 Contrast versus Spatial Resolution in Digital Imaging 315

Chapter 12: Adjuncts to Radiology 317

- 12.1 Geometric Tomography 317
- 12.2 Digital Tomosynthesis 320
- 12.3 Temporal Subtraction 321
- 12.4 Dual-Energy Subtraction 323

Chapter 13: Computed Tomography 327

- 13.1 Basic Principles 327
- 13.2 Geometry and Historical Development 331
- 13.3 Detectors and Detector Arrays 339
- 13.4 Details of Acquisition 342
- 13.5 Tomographic Reconstruction 346
- 13.6 Digital Image Display 358
- 13.7 Radiation Dose 362
- 13.8 Image Quality 367
- 13.9 Artifacts 369

Chapter 14: Nuclear Magnetic Resonance 373

- 14.1 Magnetization Properties 373
- 14.2 Generation and Detection of the Magnetic Resonance Signal 381
- 14.3 Pulse Sequences 391
- 14.4 Spin Echo 391
- 14.5 Inversion Recovery 399
- 14.6 Gradient Recalled Echo 403
- 14.7 Signal from Flow 408
- 14.8 Perfusion and Diffusion Contrast 409
- 14.9 Magnetization Transfer Contrast 411

Chapter 15: Magnetic Resonance Imaging (MRI) 415

- 15.1 Localization of the MR Signal 415
- 15.2 k-space Data Acquisition and Image Reconstruction 426
- 15.3 Three-Dimensional Fourier Transform Image Acquisition 438
- 15.4 Image Characteristics 439
- 15.5 Angiography and Magnetization Transfer Contrast 442
- 15.6 Artifacts 447
- 15.7 Instrumentation 458
- 15.8 Safety and Bioeffects 465

Chapter 16: Ultrasound 469

- 16.1 Characteristics of Sound 470
- 16.2 Interactions of Ultrasound with Matter 476

16.3	Transducers	483
16.4	Beam Properties	490
16.5	Image Data Acquisition	501
16.6	Two-Dimensional Image Display and Storage	510
16.7	Miscellaneous Issues	516
16.8	Image Quality and Artifacts	524
16.9	Doppler Ultrasound	531
16.10	System Performance and Quality Assurance	544
16.11	Acoustic Power and Bioeffects	548
Chapter 17:	Computer Networks, PACS, and Teleradiology	555
17.1	Computer Networks	555
17.2	PACS and Teleradiology	565
SECTION III: NUCLEAR MEDICINE 587		
Chapter 18:	Radioactivity and Nuclear Transformation	589
18.1	Radionuclide Decay Terms and Relationships	589
18.2	Nuclear Transformation	593
Chapter 19:	Radionuclide Production and Radiopharmaceuticals	603
19.1	Radionuclide Production	603
19.2	Radiopharmaceuticals	617
19.3	Regulatory Issues	624
Chapter 20:	Radiation Detection and Measurement	627
20.1	Types of Detectors	627
20.2	Gas-Filled Detectors	632
20.3	Scintillation Detectors	636
20.4	Semiconductor Detectors	641
20.5	Pulse Height Spectroscopy	644
20.6	Non-Imaging Detector Applications	654
20.7	Counting Statistics	661
Chapter 21:	Nuclear Imaging—The Scintillation Camera	669
21.1	Planar Nuclear Imaging: The Anger Scintillation Camera	670
21.2	Computers in Nuclear Imaging	695

Chapter 22: Nuclear Imaging—Emission Tomography 703

- 22.1 Single Photon Emission Computed Tomography (SPECT) 704
- 22.2 Positron Emission Tomography (PET) 719

SECTION IV: RADIATION PROTECTION, DOSIMETRY, AND BIOLOGY 737

Chapter 23: Radiation Protection 739

- 23.1 Sources of Exposure to Ionizing Radiation 739
- 23.2 Personnel Dosimetry 747
- 23.3 Radiation Detection Equipment in Radiation Safety 753
- 23.4 Radiation Protection and Exposure Control 755
- 23.5 Regulatory Agencies and Radiation Exposure Limits 788

Chapter 24: Radiation Dosimetry of the Patient 795

- 24.1 X-ray Dosimetry 800
- 24.2 Radiopharmaceutical Dosimetry: The MIRD Method 805

Chapter 25: Radiation Biology 813

- 25.1 Interaction of Radiation with Tissue 814
- 25.2 Cellular Radiobiology 818
- 25.3 Response of Organ Systems to Radiation 827
- 25.4 Acute Radiation Syndrome 831
- 25.5 Radiation-Induced Carcinogenesis 838
- 25.6 Hereditary Effects of Radiation Exposure 851
- 25.7 Radiation Effects *In Utero* 853

SECTION V: APPENDICES 863

Appendix A: Fundamental Principles of Physics 865

- A.1 Physical Laws, Quantities, and Units 865
- A.2 Classical Physics 867
- A.3 Electricity and Magnetism 868

Appendix B: Physical Constants, Prefixes, Geometry, Conversion Factors, and Radiologic Data 883

B.3 Radiological Data for Elements 1 through 100 885

Appendix C: Mass Attenuation Coefficients and Spectra Data Tables 887

C.1 Mass Attenuation Coefficients for Selected Elements 887

C.2 Mass Attenuation Coefficients for Selected Compounds 889

C.3 Mass Energy Attenuation Coefficients for Selected Detector Compounds 890

C.4 Mammography Spectra: Mo/Mo 891

C.5 Mammography Spectra: Mo/Rh 893

C.6 Mammography Spectra: Rh/Rh 895

C.7 General Diagnostic Spectra: W/Al 897

Appendix D: Radiopharmaceutical Characteristics and Dosimetry 899

D.1 Route of administration, localization, clinical utility, and other characteristics of commonly used radiopharmaceuticals 900

D.2 Typical administered adult activity, highest organ dose, gonadal dose, and adult effective dose for commonly used radiopharmaceuticals 908

D.3 Effective doses per unit activity administered to patients age 15, 10, 5, and 1 year for commonly used diagnostic radiopharmaceuticals 910

D.4 Absorbed dose estimates to the embryo/fetus per unit activity administered to the mother for commonly used radiopharmaceuticals 911

Appendix E: Internet Resources 913

Subject Index 915

S E C T I O N

I

BASIC CONCEPTS

INTRODUCTION TO MEDICAL IMAGING

Medical imaging of the human body requires some form of energy. In the medical imaging techniques used in radiology, the energy used to produce the image must be capable of penetrating tissues. Visible light, which has limited ability to penetrate tissues at depth, is used mostly outside of the radiology department for medical imaging. Visible light images are used in dermatology (skin photography), gastroenterology and obstetrics (endoscopy), and pathology (light microscopy). Of course, all disciplines in medicine use direct visual observation, which also utilizes visible light. In diagnostic radiology, the electromagnetic spectrum outside the visible light region is used for x-ray imaging, including mammography and computed tomography, magnetic resonance imaging, and nuclear medicine. Mechanical energy, in the form of high-frequency sound waves, is used in ultrasound imaging.

With the exception of nuclear medicine, all medical imaging requires that the energy used to penetrate the body's tissues also interact with those tissues. If energy were to pass through the body and not experience some type of interaction (e.g., absorption, attenuation, scattering), then the detected energy would not contain any useful information regarding the internal anatomy, and thus it would not be possible to construct an image of the anatomy using that information. In nuclear medicine imaging, radioactive agents are injected or ingested, and it is the metabolic or physiologic *interactions* of the agent that give rise to the information in the images.

While medical images can have an aesthetic appearance, the diagnostic utility of a medical image relates to both the technical quality of the image and the conditions of its acquisition. Consequently, the assessment of image quality in medical imaging involves very little artistic appraisal and a great deal of technical evaluation. In most cases, the image quality that is obtained from medical imaging devices involves compromise—better x-ray images can be made when the radiation dose to the patient is high, better magnetic resonance images can be made when the image acquisition time is long, and better ultrasound images result when the ultrasound power levels are large. Of course, patient safety and comfort must be considered when acquiring medical images; thus excessive patient dose in the pursuit of a perfect image is not acceptable. Rather, the power levels used to make medical images require a balance between patient safety and image quality.

1.1 THE MODALITIES

Different types of medical images can be made by varying the types of energies used and the acquisition technology. The different modes of making images are referred to as *modalities*. Each modality has its own applications in medicine.

Radiography

Radiography was the first medical imaging technology, made possible when the physicist Wilhelm Roentgen discovered x-rays on November 8, 1895. Roentgen also made the first radiographic images of human anatomy (Fig. 1-1). Radiography (also called roentgenography) defined the field of radiology, and gave rise to radiologists, physicians who specialize in the interpretation of medical images. Radiography is performed with an x-ray source on one side of the patient, and a (typically flat) x-ray detector on the other side. A short duration (typically less than $\frac{1}{2}$ second) pulse of x-rays is emitted by the x-ray tube, a large fraction of the x-rays interacts in the patient, and some of the x-rays pass through the patient and reach the detector, where a radiographic image is formed. The homogeneous distribution of x-rays that enter the patient is modified by the degree to which the x-rays are removed from the beam (i.e., attenuated) by scattering and absorption within the tissues. The attenuation properties of tissues such as bone, soft tissue, and air inside the patient are very different, resulting in the heterogeneous distribution of x-rays that emerges from the patient. The radiographic image is a picture of this x-ray distribution. The detector used in radiography can be photographic film (e.g., screen-film radiography) or an electronic detector system (i.e., digital radiography).



FIGURE 1-1. The beginning of diagnostic radiology, represented by this famous radiographic image made on December 22, 1895 of the wife of the discoverer of x-rays, Wilhelm Conrad Roentgen. The bones of her hand as well as two rings on her finger are clearly visible. Within a few months, Roentgen was able to determine the basic physical properties of x-rays. Nearly simultaneously, as word of the discovery spread around the world, medical applications of this "new kind of ray" propelled radiologic imaging into an essential component of medical care. In keeping with mathematical conventions, Roentgen assigned the letter "x" to represent the unknown nature of the ray and thus the term x-ray was born. Details regarding x-ray production and interactions can be found in Chapters 5 and 3, respectively. (Reproduced from Glasser O. *Wilhelm Conrad and Röntgen and the early history of the roentgen rays*. Springfield, IL: Charles C. Thomas, 1933, with permission.)



FIGURE 1-2. The chest x-ray is the most ubiquitous image in diagnostic radiology. High x-ray energy is used for the purpose of penetrating the mediastinum, cardiac, and diaphragm areas of the patient without overexposing areas within the lungs. Variation in the gray-scale image represents an attenuation map of the x-rays: dark areas (high film optical density) have low attenuation, and bright areas (low film optical density) have high attenuation. The image here shows greater than normal attenuation in the lower lobes of the lungs, consistent with plural effusion, right greater than left. Rapid acquisition, low risk, low cost, and high diagnostic value are the major reasons why x-ray projection imaging represents the bulk of all diagnostic imaging studies (typically 60% to 70% of all images produced). Projection imaging physics is covered in Chapter 6.

Transmission imaging refers to imaging in which the energy source is outside the body on one side, and the energy passes through the body and is detected on the other side of the body. Radiography is a transmission imaging modality. *Projection imaging* refers to the case when each point on the image corresponds to information along a straight line trajectory through the patient. Radiography is also a projection imaging modality. Radiographic images are useful for a very wide range of medical indications, including the diagnosis of broken bones, lung cancer, cardiovascular disorders, etc. (Fig. 1-2).

Fluoroscopy

Fluoroscopy refers to the continuous acquisition of a sequence of x-ray images over time, essentially a real-time x-ray movie of the patient. Fluoroscopy is a transmission projection imaging modality, and is, in essence, just real-time radiography. Fluoroscopic systems use x-ray detector systems capable of producing images in rapid temporal sequence. Fluoroscopy is used for positioning catheters in arteries, for visualizing contrast agents in the gastrointestinal (GI) tract, and for other medical applications such as invasive therapeutic procedures where real-time image feedback is necessary. Fluoroscopy is also used to make x-ray movies of anatomic motion, such as of the heart or the esophagus.

Mammography

Mammography is radiography of the breast, and is thus a transmission projection type of imaging. Much lower x-ray energies are used in mammography than any other radiographic applications, and consequently modern mammography uses



FIGURE 1-3. Mammography is a specialized x-ray projection imaging technique useful for detecting breast anomalies such as masses and calcifications. The mammogram in the image above demonstrates a spiculated mass (*arrow*) with an appearance that is typical of a cancerous lesion in the breast, in addition to blood vessels and normal anatomy. Dedicated mammography equipment using low x-ray energies, k-edge filters, compression, screen/film detectors, antiscatter grids, and automatic exposure control results in optimized breast images of high quality and low x-ray dose, as detailed in Chapter 8. X-ray mammography is the current procedure of choice for screening and early detection of breast cancer because of high sensitivity, excellent benefit to risk, and low cost.

x-ray machines and detector systems specifically designed for breast imaging. Mammography is used to screen asymptomatic women for breast cancer (screening mammography), and is also used to help in the diagnosis of women with breast symptoms such as the presence of a lump (diagnostic mammography) (Fig. 1-3).

Computed Tomography (CT)

CT became clinically available in the early 1970s and is the first medical imaging modality made possible by the computer. CT images are produced by passing x-rays through the body, at a large number of angles, by rotating the x-ray tube around the body. One or more linear detector arrays, opposite the x-ray source, collect the transmission projection data. The numerous data points collected in this manner are synthesized by a computer into a *tomographic* image of the patient. The term *tomography* refers to a picture (*-graph*) of a slice (*tomo-*). CT is a transmission technique that results in images of individual slabs of tissue in the patient. The advantage of a tomographic image over projection image is its ability to display the anatomy in a slab (slice) of tissue in the absence of over- or underlying structures.

CT changed the practice of medicine by substantially reducing the need for exploratory surgery. Modern CT scanners can acquire 5-mm-thick tomographic images along a 30-cm length of the patient (i.e., 60 images) in 10 seconds, and reveal the presence of cancer, ruptured discs, subdural hematomas, aneurysms, and a large number of other pathologies (Fig. 1-4).



FIGURE 1-4. A computed tomography (CT) image of the abdomen reveals a ruptured disc (arrow) manifested as the bright area of the image adjacent to the vertebral column. Anatomic structures such as the kidneys, arteries, and intestines are clearly represented in the image. CT provides high-contrast sensitivity for soft tissue, bone, and air interfaces without superimposition of anatomy. With recently implemented multiple array detectors, scan times of 0.5 seconds per 360 degrees and fast computer reconstruction permits head-to-toe imaging in as little as 30 seconds. Because of fast acquisition speed, high-contrast sensitivity, and ability to image tissue, bone, and air, CT remains the workhorse of tomographic imaging in diagnostic radiology. Chapter 13 describes the details of CT.

Nuclear Medicine Imaging

Nuclear medicine is the branch of radiology in which a chemical or compound containing a radioactive isotope is given to the patient orally, by injection, or by inhalation. Once the compound has distributed itself according to the physiologic status of the patient, a radiation detector is used to make projection images from the x- and/or gamma rays emitted during radioactive decay of the agent. Nuclear medicine produces *emission* images (as opposed to transmission images), because the radioisotopes emit their energy from inside the patient.

Nuclear medicine imaging is a form of *functional* imaging. Rather than yielding information about just the anatomy of the patient, nuclear medicine images provide information regarding the physiologic conditions in the patient. For example, thallium tends to concentrate in normal heart muscle, but does not concentrate as well in areas that are infarcted or ischemic. These areas appear as “cold spots” on a nuclear medicine image, and are indicative of the functional status of the heart. Thyroid tissue has a great affinity for iodine, and by administering radioactive iodine (or its analogs), the thyroid can be imaged. If thyroid cancer has metastasized in the patient, then “hot spots” indicating their location will be present on the nuclear medicine images. Thus functional imaging is the forte of nuclear medicine.

Nuclear Medicine Planar Imaging

Nuclear medicine planar images are projection images, since each point on the image is representative of the radioisotope activity along a line projected through the patient. Planar nuclear images are essentially two-dimensional maps of the radioisotope distribution, and are helpful in the evaluation of a large number of disorders (Fig. 1-5).

Single Photon Emission Computed Tomography (SPECT)

SPECT is the tomographic counterpart of nuclear medicine planar imaging, just like CT is the tomographic counterpart of radiography. In SPECT, a nuclear camera records x- or gamma-ray emissions from the patient from multiple angles, and a computer reconstructs a three-dimensional image.



FIGURE 1-5. Anterior and posterior whole-body bone scan of a 74-year-old woman with a history of right breast cancer. This patient was injected with 925 MBq (25 mCi) of technetium (Tc) 99m methylenediphosphonate (MDP) and was imaged 3 hours later with a dual-headed whole-body scintillation camera. The scan demonstrates multiple areas of osteoblastic metastases in the axial and proximal skeleton. Incidental findings include an arthritis pattern in the shoulders and left knee. Computer processed planar imaging is still the standard for many nuclear medicine examinations (e.g., whole-body bone scans, hepatobiliary, thyroid, renal, and pulmonary studies). Planar nuclear imaging is discussed in detail in Chapter 21. (Image courtesy of Dr. David Shelton, University of California, Davis Medical Center, Davis, CA.)



FIGURE 1-6. A myocardial perfusion stress test utilizing thallium 201 (Tl 201) and single photon emission computed tomography (SPECT) imaging was performed on a 79-year-old woman with chest pain. This patient had pharmacologic stress with dipyridamole and was injected with 111 MBq (3 mCi) of Tl 201 at peak stress. Stress imaging followed immediately on a variable-angle two-headed SPECT camera. Image data was acquired over 180 degrees at 30 seconds per stop. The rest/redistribution was done 3 hours later with a 37-MBq (1-mCi) booster injection of Tl 201. Short axis, horizontal long axis, and vertical long axis views show relatively reduced perfusion on anterolateral wall stress images, with complete reversibility on rest/redistribution images. Findings indicated coronary stenosis in the left anterior descending (LAD) coronary artery distribution. SPECT is now the standard for a number of nuclear medicine examinations including cardiac perfusion and brain and tumor imaging. SPECT imaging is discussed in detail in Chapter 22. (Image courtesy of Dr. David Shelton, University of California, Davis Medical Center, Davis, CA.)

angles around the patient. These projection data are used to reconstruct a series of tomographic emission images. SPECT images provide diagnostic functional information similar to nuclear planar examinations; however, their tomographic nature allows physicians to better understand the precise distribution of the radioactive agent, and to make a better assessment of the function of specific organs or tissues within the body (Fig. 1-6). The same radioactive isotopes are used in both planar nuclear imaging and SPECT.

Positron Emission Tomography (PET)

Positrons are positively charged electrons, and are emitted by some radioactive isotopes such as fluorine 18 and oxygen 15. These radioisotopes are incorporated into metabolically relevant compounds [such as ^{18}F -fluorodeoxyglucose (FDG)], which localize in the body after administration. The decay of the isotope produces a positron, which rapidly undergoes a very unique interaction: the positron (e^+) combines with an electron (e^-) from the surrounding tissue, and the mass of both the e^+ and the e^- is converted by annihilation into pure energy, following Einstein's famous equation $E = mc^2$. The energy that is emitted is called *annihilation radiation*. Annihilation radiation production is similar to gamma-ray emission, except that *two* photons are emitted, and they are emitted in almost exactly opposite directions, i.e., 180 degrees from each other. A PET scanner utilizes a ring of detectors that surround the patient, and has special circuitry that is capable of identifying the photon pairs produced during annihilation. When a photon pair is detected by two detectors on the scanner, it is known that the decay event took place somewhere along a straight line between those two detectors. This information is used to mathematically compute the three-dimensional distribution of the PET agent, resulting in a series of tomographic emission images.

Although more expensive than SPECT, PET has clinical advantages in certain diagnostic areas. The PET detector system is more sensitive to the presence of radioisotopes than SPECT cameras, and thus can detect very subtle pathologies. Furthermore, many of the elements that emit positrons (carbon, oxygen, fluorine) are quite physiologically relevant (fluorine is a good substitute for a hydroxyl group), and can be incorporated into a large number of biochemicals. The most important of these is ^{18}F FDG, which is concentrated in tissues of high glucose metabolism such as primary tumors and their metastases (Fig. 1-7).

Magnetic Resonance Imaging (MRI)

MRI scanners use magnetic fields that are about 10,000 to 60,000 times stronger than the earth's magnetic field. Most MRI utilizes the nuclear magnetic resonance properties of the proton—i.e., the nucleus of the hydrogen atom, which is very abundant in biologic tissues (each cubic millimeter of tissue contains about 10^{18} protons). The proton has a magnetic moment, and when placed in a 1.5-tesla (T) magnetic field, the proton will preferentially absorb radio wave energy at the resonance frequency of 63 megahertz (MHz).

In MRI, the patient is placed in the magnetic field, and a pulse of radio waves is generated by antennas ("coils") positioned around the patient. The protons in the patient absorb the radio waves, and subsequently reemit this radio wave energy after a period of time that depends on the very localized magnetic properties of the sur-



FIGURE 1-7. Whole-body positron emission tomography (PET) scan of a 54-year-old man with malignant melanoma. Patient was injected intravenously with 600 MBq (16 mCi) of ^{18}F -deoxyglucose. The patient was imaged for 45 minutes, beginning 75 minutes after injection of the radiopharmaceutical. The image demonstrates extensive metastatic disease with abnormalities throughout the axial and proximal appendicular skeleton, right and left lungs, liver, and left inguinal and femoral lymph nodes. The unique ability of the PET scan in this case was to correctly assess the extent of disease, which was underestimated by CT, and to serve as a baseline against which future comparisons could be made to assess the effects of immunochemotherapy. PET has applications in functional brain and cardiac imaging and is rapidly becoming a routine diagnostic tool in the staging of many cancers. PET technology is discussed in detail in Chapter 22. (Image courtesy of Dr. George Segall, VA Medical Center, Palo Alto, CA.)

rounding tissue. The radio waves emitted by the protons in the patient are detected by the antennas that surround the patient. By slightly changing the strength of the magnetic field as a function of position in the patient (using magnetic field *gradients*), the proton resonance frequency will vary as a function of position, since frequency is proportional to magnetic field strength. The MRI system uses the frequency (and phase) of the returning radio waves to determine the position of each signal from the patient. The mode of operation of MRI systems is often referred to as *spin echo* imaging.

MRI produces a set of tomographic slices through the patient, where each point in the image depends on the micromagnetic properties of the corresponding tissue at that point. Because different types of tissue such as fat, white, and gray matter in the brain, cerebral spinal fluid, and cancer all have different local magnetic properties, images made using MRI demonstrate high sensitivity to anatomic variations and therefore are high in contrast. MRI has demonstrated exceptional utility in neurologic imaging (head and spine), and for musculoskeletal applications such as imaging the knee after athletic injury (Fig. 1-8).

MRI is a tomographic imaging modality, and competes with x-ray CT in many clinical applications. The acquisition of the highest-quality images using MRI requires tens of minutes, whereas a CT scan of the entire head requires about 10 seconds. Thus, for patients where motion cannot be controlled (pediatric patients) or in anatomic areas where involuntary patient motion occurs (the beating heart

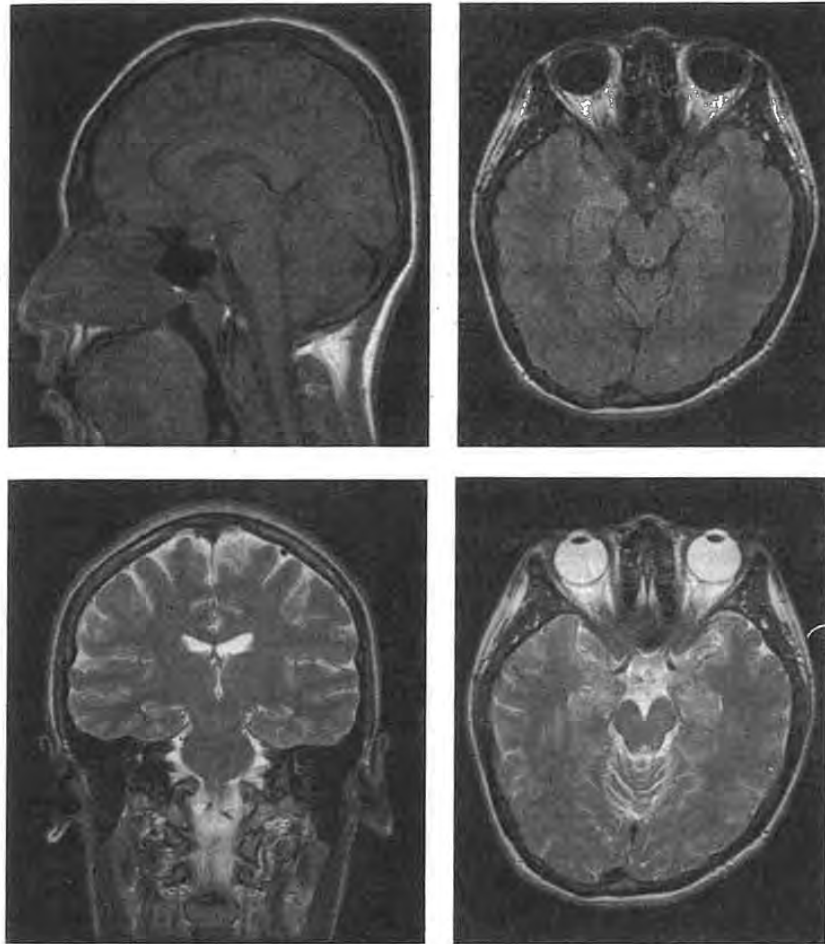


FIGURE 1-8. Sagittal (*upper left*), coronal (*lower left*), and axial (*right*) normal images of the brain demonstrate the exquisite contrast sensitivity and acquisition capability of magnetic resonance imaging (MRI). Image contrast is generated by specific MR pulse sequences (*upper images* are T1-weighted, *lower images* are T2 weighted) to emphasize the magnetization characteristics of the tissues placed in a strong magnetic field, based on selective excitation and reemission of radiofrequency signals. MRI is widely used for anatomic as well as physiologic and functional imaging, and is the modality of choice for neurologic assessment, staging of cancer, and physiologic/anatomic evaluation of extremities and joints. Further information regarding MRI can be found in Chapters 14 and 15.

and the churning intestines), CT is often used instead of MRI. Also, because of the large magnetic field used in MRI, electronic monitoring equipment cannot be used while the patient is being scanned. Thus, for most trauma, CT is preferred. MRI should not be performed on patients who have cardiac pacemakers or internal ferromagnetic objects such as surgical aneurysm clips, metal plate or rod implants, or metal shards near critical structures like the eye.

Despite some indications where MRI should not be used, fast image acquisition techniques using special coils have made it possible to produce images in much shorter periods of time, and this has opened up the potential of using MRI for imaging of the motion-prone thorax and abdomen. MRI scanners can also detect the presence of motion, which is useful for monitoring blood flow through arteries (MR *angiography*).

Ultrasound Imaging

When a book is dropped on a table, the impact causes pressure waves (called sound) to propagate through the air such that they can be heard at a distance. Mechanical energy in the form of high-frequency (“ultra”) sound can be used to generate images of the anatomy of a patient. A short-duration pulse of sound is generated by an ultrasound *transducer* that is in direct physical contact with the tissues being imaged. The sound waves travel into the tissue, and are reflected by internal structures in the body, creating echoes. The reflected sound waves then reach the trans-



FIGURE 1-9. The ultrasound image is a map of the echoes from tissue boundaries of high-frequency sound wave pulses (typically from 2 to 20 MHz frequency) gray-scale encoded into a two-dimensional tomographic image. A phased-array transducer operating at 3.5 MHz produced this normal obstetrical ultrasound image of Jennifer Lauren Bushberg (at approximately 3½ months). Variations in the image brightness are due to acoustic characteristics of the tissues; for example, the fluid in the placenta is echo free, while most fetal tissues are echogenic and produce a larger signal strength. Shadowing is caused by highly attenuating or scattering tissues such as bone or air and the corresponding distal low-intensity streaks. Besides tomographic acoustic imaging, distance measurements (e.g., fetal head diameter assessment for aging), and vascular assessment using Doppler techniques and color flow imaging are common. Increased use of ultrasound is due to low equipment cost, portability, high safety, and minimal risk. Details of ultrasound are found in Chapter 16.

ducer, which records the returning sound beam. This mode of operation of an ultrasound device is called *pulse echo* imaging. The sound beam is swept over a range of angles (a sector) and the echoes from each line are recorded and used to compute an ultrasonic image in the shape of a sector (*sector scanning*).

Ultrasound is reflected strongly by interfaces, such as the surfaces and internal structures of abdominal organs. Because ultrasound is less harmful than ionizing radiation to a growing fetus, ultrasound imaging is preferred in obstetric patients (Fig. 1-9). An interface between tissue and air is highly echoic, and thus very little sound can penetrate from tissue into an air-filled cavity. Therefore, ultrasound imaging has less utility in the thorax where the air in the lungs presents a wall that the sound beam cannot penetrate. Similarly, an interface between tissue and bone is also highly echoic, thus making brain imaging, for example, impractical in most cases.

Doppler Ultrasound Imaging

Doppler imaging using ultrasound makes use of a phenomenon familiar to train enthusiasts. For the observer standing beside railroad tracks as a rapidly moving train goes by blowing its whistle, the pitch of the whistle is higher as the train approaches and becomes lower as the train passes by the observer and speeds off into the distance. The change in the pitch of the whistle, which is an apparent change in the frequency of the sound, is a result of the Doppler effect. The same phenomenon occurs at ultrasound frequencies, and the change in frequency (the Doppler shift) is used to measure the motion of blood or of the heart. Both the velocity and direction of blood flow can be measured, and color Doppler display usually shows blood flow in one direction as red and in the other direction as blue.

1.2 IMAGE PROPERTIES

Contrast

Contrast in an image is the difference in the gray scale of the image. A uniformly gray image has no contrast, whereas an image with vivid transitions between dark gray and light gray demonstrates high contrast. Each imaging modality generates contrast based on different physical parameters in the patient.

X-ray contrast (radiography, fluoroscopy, mammography, and CT) is produced by differences in tissue composition, which affect the local x-ray absorption coefficient, which in turn is dependent on the density (g/cm^3) and the effective atomic number. The energy of the x-ray beam (adjusted by the operator) also affects contrast in x-ray images. Because bone has a markedly different effective atomic number ($Z_{\text{eff}} \approx 13$) than soft tissue ($Z_{\text{eff}} \approx 7$), due to its high concentration of calcium ($Z = 20$) and phosphorus ($Z = 15$), bones produce high contrast on x-ray image modalities. The chest radiograph, which demonstrates the lung parenchyma with high tissue/airway contrast, is the most common radiographic procedure performed in the world (Fig. 1-2).

CT's contrast is enhanced over other radiographic imaging modalities due to its tomographic nature. The absence of out-of-slice structures in the CT image greatly improves its image contrast.

Nuclear medicine images (planar images, SPECT, and PET) are maps of the spatial distribution of radioisotopes in the patients. Thus, contrast in nuclear images depends on the tissue's ability to concentrate the radioactive material. The uptake of a radiopharmaceutical administered to the patient is dependent on the pharmacologic interaction of the agent with the body. PET and SPECT have much better contrast than planar nuclear imaging because, like CT, the images are not obscured by out-of-slice structures.

Contrast in MRI is related primarily to the proton density and to relaxation phenomena (i.e., how fast a group of protons gives up its absorbed energy). Proton density is influenced by the mass density (g/cm^3), so MRI can produce images that look somewhat like CT images. Proton density differs among tissue types, and in particular adipose tissues have a higher proportion of protons than other tissues, due to the high concentration of hydrogen in fat [$\text{CH}_3(\text{CH}_2)_n\text{COOH}$]. Two different relaxation mechanisms (spin/lattice and spin/spin) are present in tissue, and the dominance of one over the other can be manipulated by the timing of the radiofrequency (RF) pulse sequence and magnetic field variations in the MRI system. Through the clever application of different pulse sequences, blood flow can be detected using MRI techniques, giving rise to the field of MR angiography. Contrast mechanisms in MRI are complex, and thus provide for the flexibility and utility of MR as a diagnostic tool.

Contrast in ultrasound imaging is largely determined by the acoustic properties of the tissues being imaged. The difference between the *acoustic impedances* (tissue density $\times$ speed of sound in tissue) of two adjacent tissues or other substances affects the amplitude of the returning ultrasound signal. Hence, contrast is quite apparent at tissue interfaces where the differences in acoustic impedance are large. Thus, ultrasound images display unique information about patient anatomy not provided by other imaging modalities. Doppler ultrasound imaging shows the amplitude and direction of blood flow by analyzing the frequency shift in the reflected signal, and thus motion is the source of contrast.

Spatial Resolution

Just as each modality has different mechanisms for providing contrast, each modality also has different abilities to resolve fine detail in the patient. *Spatial resolution* refers to the ability to see small detail, and an imaging system has *higher* spatial resolution if it can demonstrate the presence of *smaller* objects in the image. The *limiting spatial resolution* is the size of the smallest object that an imaging system can resolve.

Table 1-1 lists the limiting spatial resolution of each imaging modality used in medical imaging. The wavelength of the energy used to probe the object is a fundamental limitation of the spatial resolution of an imaging modality. For example, optical microscopes cannot resolve objects smaller than the wavelengths of visible light, about 500 nm. The wavelength of x-rays depends on the x-ray energy, but even the longest x-ray wavelengths are tiny—about one ten-billionth of a meter. *Table 1-1. Limiting spatial resolution of imaging modalities.* This table represents the theoretical limiting spatial resolution of each imaging modality. Actual resolution is often lower due to various factors.

TABLE 1-1. THE LIMITING SPATIAL RESOLUTIONS OF VARIOUS MEDICAL IMAGING MODALITIES: THE RESOLUTION LEVELS ACHIEVED IN TYPICAL CLINICAL USAGE OF THE MODALITY

Modality	Δ (mm)	Comments
Screen film radiography	0.08	Limited by focal spot and detector resolution
Digital radiography	0.17	Limited by size of detector elements
Fluoroscopy	0.125	Limited by detector and focal spot
Screen film mammography	0.03	Highest resolution modality in radiology
Digital mammography	0.05–0.10	Limited by size of detector elements
Computed tomography	0.4	About 1/2-mm pixels
Nuclear medicine planar imaging	7	Spatial resolution degrades substantially with distance from detector
Single photon emission computed tomography	7	Spatial resolution worst toward the center of cross-sectional image slice
Positron emission tomography	5	Better spatial resolution than with the other nuclear imaging modalities
Magnetic resonance imaging	1.0	Resolution can improve at higher magnetic fields
Ultrasound imaging (5 MHz)	0.3	Limited by wavelength of sound

MRI poses a paradox to the wavelength imposed resolution rule—the wavelength of the RF waves used (at 1.5 T, 63 MHz) is 470 cm, but the spatial resolution of MRI is better than a millimeter. This is because the spatial distribution of the paths of RF energy is not used to form the actual image (contrary to ultrasound, light microscopy, and x-ray images). The RF energy is collected by a large antenna, and it carries the spatial information of a group of protons encoded in its frequency spectrum.

COMPUTERS IN MEDICAL IMAGING

Computers were originally designed to perform mathematical computations and other data processing tasks very quickly. Since then, they have come to be used for many other purposes, including information display, information storage, and, in conjunction with computer networks, information transfer and communications.

Computers were introduced in medical imaging in the early 1970s and have become increasingly important since that time. Their first uses in medical imaging were mostly in nuclear medicine, where they were used to acquire series of images depicting the organ-specific kinetics of radiopharmaceuticals. Today, computers are essential to many imaging modalities, including x-ray computed tomography (CT), magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT), positron emission tomography (PET), and digital radiography.

Any function that can be performed by a computer can also be performed by a hard-wired electronic circuit. The advantage of the computer over a hard-wired circuit is its flexibility. The function of the computer can be modified merely by changing the program that controls the computer, whereas modifying the function of a hard-wired circuit usually requires replacing the circuit. Although the computer is a very complicated and powerful data processing device, the actual operations performed by a computer are very simple. The power of the computer is mainly due to its speed in performing these operations and its ability to store large volumes of data.

In the next section of this chapter, the storage and transfer of data in digital form are discussed. The following sections describe the components of a computer; factors that determine the performance of a computer; computer software; and the acquisition, storage, and display of digital images. Computer networks, picture archiving, and communications systems (PACS), and teleradiology are discussed in detail in Chapter 17.

4.1 STORAGE AND TRANSFER OF DATA IN COMPUTERS

Number Systems

Our culture uses a number system based on ten, probably because humans have five fingers on each hand and number systems having evolved from the simple act of counting on the fingers. Computers use the binary system for the electronic storage and manipulation of numbers.

Decimal Form (Base 10)

In the decimal form, the ten digits 0 through 9 are used to represent numbers. To represent numbers greater than 9, several digits are placed in a row. The value of each digit in a number depends on its position in the row; the value of a digit in any position is *ten* times the value it would represent if it were shifted one place to the right. For example, the decimal number 3,506 actually represents:

$$(3 \times 10^3) + (5 \times 10^2) + (0 \times 10^1) + (6 \times 10^0)$$

where $10^1 = 10$ and $10^0 = 1$. The leftmost digit in a number is called the *most significant digit* and the rightmost digit is called the *least significant digit*.

Binary Form (Base 2)

In binary form, the two digits 0 and 1 are used to represent numbers. Each of these two digits, by itself, has the same meaning that it has in the decimal form. To represent numbers greater than 1, several digits are placed in a row. The value of each digit in a number depends on its position in the row; the value of a digit in any position is *two* times the value it would represent if it were shifted one place to the right. For example, the binary number 1101 represents:

$$(1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0)$$

where $2^3 = 8$, $2^2 = 4$, $2^1 = 2$, and $2^0 = 1$. To count in binary form, 1 is added to the least significant digit of a number. If the least significant digit is 1, it is replaced by 0 and 1 is added to the next more significant digit. If several contiguous digits on the right are 1, each is replaced by 0 and 1 is added to the least significant digit that was not 1. Counting the binary system is illustrated in Table 4-1.

Conversions Between Decimal and Binary Forms

To convert a number from binary to decimal form, the binary number is expressed as a series of powers of two and the terms in the series are added. For example, to convert the binary number 101010 to decimal form:

$$101010 \text{ (binary)} = (1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (0 \times 2^0)$$

TABLE 4-1. NUMBERS IN DECIMAL AND BINARY FORMS

Decimal	Binary	Decimal	Binary
0	0	8	1000
1	1	9	1001
2	10	10	1010
3	11	11	1011
4	100	12	1100
5	101	13	1101
6	110	14	1110
7	111	15	1111
		16	10000

TABLE 4-2. CONVERSION OF 42 (DECIMAL) INTO BINARY FORM^a

Division	Result	Remainder	
42/2 =	21	0	Least significant digit
21/2 =	10	1	
10/2 =	5	0	
5/2 =	2	1	
2/2 =	1	0	
1/2 =	0	1	Most significant digit

^aIt is repeatedly divided by 2, with the remainder recorded after each division. The binary equivalent of 42 (decimal) is therefore 101010.

$$= (1 \times 32) + (1 \times 8) + (1 \times 2) = 42 \text{ (decimal)}.$$

To convert a number from decimal into binary representation, it is repeatedly divided by two. Each division determines one digit of the binary representation, starting with the least significant. If there is no remainder from the division, the digit is 0; if the remainder is 1, the digit is 1. The conversion of 42 (decimal) into binary form is illustrated in Table 4-2.

Considerations Regarding Number Systems

Whenever it is not clear which form is being used, the form will be written in parentheses after the number. If the form is not specified, it can usually be assumed that a number is in decimal form. It is important not to confuse the binary representation of a number with its more familiar decimal representation. For example, 10 (binary) and 10 (decimal) represent different numbers, although they look alike. On the other hand, 1010 (binary) and 10 (decimal) represent the same number. The only numbers that appear the same in both systems are 0 and 1.

Digital Representation of Data

Bits, Bytes, and Words

Computer memory and storage consist of many elements called *bits* (for *binary digits*), each of which can be in one of two states. A bit is similar to a light switch, because a light switch can only be on or off. In most computer memory, each bit is a transistor; *on* is one state and *off* is the other. In magnetic storage media, such as magnetic disks and magnetic tape, a bit is a small portion of the disk or tape that may be magnetized.

Bits are grouped into *bytes*, each consisting of eight bits. The capacity of a computer memory or storage device is usually described in kilobytes, megabytes, gigabytes, or terabytes (Table 4-3). Bits are also grouped into *words*. The number of bits in a word depends on the computer system; 16-, 32-, and 64-bit words are common.

Digital Representation of Different Types of Data

General-purpose computers must be able to store and process several types of data in digital form. For example, if a computer is to execute a word processing program

TABLE 4-3. UNITS TO DESCRIBE COMPUTER STORAGE CAPACITY^a

1 kilobyte (kB) = 2^{10} bytes = 1024 bytes = a thousand bytes
1 megabyte (MB) = 2^{20} bytes = 1024 kilobytes = 1,048,576 bytes = a million bytes
1 gigabyte (GB) = 2^{30} bytes = 1024 megabytes = 1,073,741,824 bytes = a billion bytes
1 terabyte (TB) = 2^{40} bytes = 1024 gigabytes = 1,099,511,627,776 = a trillion bytes

^aNote that the prefixes *kilo-*, *mega-*, *giga-*, and *tera-* have slightly different meanings when used to describe computer storage capacity than in standard scientific usage.

or a program that stores and retrieves patient data, it must be able to represent alphanumeric data (text), such as a patient's name, in digital form. Computers must also be able to represent numbers in digital form. Most computers have several data formats for numbers. For example, one computer system provides formats for 1-, 2-, 4-, and 8-byte positive integers; 1-, 2-, 4-, and 8-byte signed integers ("signed" meaning that they may have positive or negative values); and 4- and 8-byte floating-point numbers (to be discussed shortly).

When numbers can assume very large or very small values or must be represented with great accuracy, formats requiring a large amount of storage per number must be used. However, if numbers are restricted to integers within a limited range of values, considerable savings in storage and processing time may be realized. For instance, the gray-scale values in ultrasound images typically range from 0 (black) to 255 (white), a range of 256 numbers. As will be seen in the following section, any gray-scale value within this range can be represented by only eight bits. The following sections describe schemes for the digital representation of different types of data.

Storage of Positive Integers

As discussed above, computer memory and storage consist of many bits. Each bit can be in one of two states and can therefore represent the numbers 0 or 1. Two bits have four possible configurations (00, 01, 10, or 11) that in decimal form are 0, 1, 2, and 3. Three bits have eight possible configurations (000, 001, 010, 011, 100, 101, 110, or 111) that can represent the decimal numbers 0, 1, 2, 3, 4, 5, 6, and 7. In general, N bits have 2^N possible configurations and can represent integers from 0 to $2^N - 1$. One byte can therefore store integers from 0 to 255 and a 16-bit word can store integers from 0 to 65,535 (Table 4-4).

TABLE 4-4. NUMBER OF BITS REQUIRED TO STORE INTEGERS

Number of Bits	Possible Configurations	Number of Configurations	Represent Integers (Decimal Form)
1	0,1	2	0,1
2	00,01,10,11	4	0,1,2,3
3	000,001,010,011,100,101,110,111	8	0,1,2,3,4,5,6,7
8	00000000 to 11111111	256	0 to 255
16	0000000000000000 to 1111111111111111	65,536	0 to 65,535
N		2^N	0 to $2^N - 1$

Binary Representation of Signed Integers

The previous discussion dealt only with positive integers. It is often necessary for computers to manipulate integers that can have positive or negative values. There are many ways to represent signed integers in binary form. The simplest method is to reserve the first bit of a number for the sign of the number. Setting the first bit to 0 can indicate that the number is positive, whereas setting it to 1 indicates that the number is negative. For an 8-bit number, 11111111 (binary) = -127 (decimal) is the most negative number and 01111111 (binary) = 127 (decimal) is the largest positive number. Most computers use a different scheme called "twos complement notation" to represent signed integers, which simplifies the circuitry needed to add positive and negative integers.

Floating-Point Form

Computers used for scientific purposes must be able to manipulate very large numbers, such as Avogadro's number (6.022×10^{23} molecules per gram-mole) and very small numbers, such as the mass of a proton (1.673×10^{-27} kilograms). These numbers are usually represented in floating-point form. Floating-point form is similar to scientific notation, in which a number is expressed as a decimal fraction times ten raised to a power. A number can be also written as a binary fraction times two to a power. For example, Avogadro's number can be written as

$$0.11111111 \text{ (binary)} \times 2^{01001111 \text{ (binary)}}$$

When a computer stores the number in floating-point form, it stores the pair of signed binary integers, 11111111 and 01001111.

Binary Representation of Text

It is often necessary for computers to store and manipulate alphanumeric data, such as a patient's name or the text of this book. One common method for representing alphanumeric data in binary form is the American Standard Code for Information Interchange (ASCII). Each character is stored in one byte. The byte values from 00000000 to 01111111 (binary) are used to define 128 characters, including upper- and lowercase letters, the digits 0 through 9, many punctuation marks, and several carriage-control characters such as line feed. For example, the carriage-control character "carriage return" is represented by 00001101, the uppercase letter "A" is represented by 01000001, the comma is represented by 00111010, and the digit "2" is represented by 00110010. Several other methods for representing text in binary form that permit more than 128 characters to be represented are also in use.

Transfers of Data in Digital Form

Data are transferred between the various components of a computer, such as the memory and central processing unit, in binary form. A voltage of fixed value (such as 5 V) and fixed duration on a wire can be used to represent the binary number 1 and a voltage of 0 V for the same duration can represent 0. A group of such voltage pulses can be used to represent a number, an alphanumeric character, or other unit of information. (This scheme is called "unipolar" digital encoding. Many other digital encoding schemes exist.)

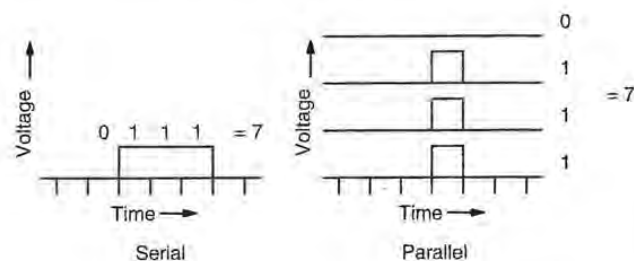


FIGURE 4-1. Serial (left) and parallel (right) transfers of digital data.

A unit of data (e.g., a byte or 16-bit word) may be sent in *serial* form over a single wire by transferring each bit value in succession, as shown on the left in Figure 4-1. A much faster way of transferring the same unit of data is to use several wires, with each bit value being transferred over its own wire, as shown on the right in Fig. 4-1. This is called *parallel* data transfer. If eight wires are used for data transfer, a byte can be transferred in an eighth of the time required for serial transfer.

A group of wires used to transfer data between several devices is called a *data bus*. Each device connected to the bus is identified by an address or a range of addresses. Only one device at a time can transmit information on a bus, and in most cases only one device receives the information. The sending device transmits both the information and the address of the device that is intended to receive the information. Some buses use separate wires for the data and addresses, whereas others transmit both over the same wires. The width of a bus refers to the number of wires used to transmit data. For example, a 32-bit bus transmits 32 bits of data simultaneously. Buses usually have additional wires, such as a ground wire to establish a reference potential for the voltage signals on the other wires, wires to carry control signals, and wires at one or more constant voltages to provide electrical power to circuit boards attached to the bus.

4.2 ANALOG DATA AND CONVERSION BETWEEN ANALOG AND DIGITAL FORMS

Analog Representation of Data

Numerical data can be represented in electronic circuits by analog form instead of digital form by a voltage or voltage pulse whose amplitude is proportional to the number being represented, as shown in Fig. 4-2A. An example of analog representation is a voltage pulse produced by a photomultiplier tube attached to a scintillation detector. The amplitude (peak voltage) of the pulse is proportional to the amount of energy deposited in the detector by an x- or gamma ray. Another example is the signal from the video camera attached to the image intensifier tube of a fluoroscopy system; the voltage at each point in time is proportional to the intensity of the x-rays incident on a portion of the input phosphor of the image intensifier tube (Fig. 4-2C).

Advantages and Disadvantages of the Analog and Digital Forms

There is a major disadvantage to the electronic transmission of data in analog form—the signals become distorted. Causes of this distortion include inaccuracies

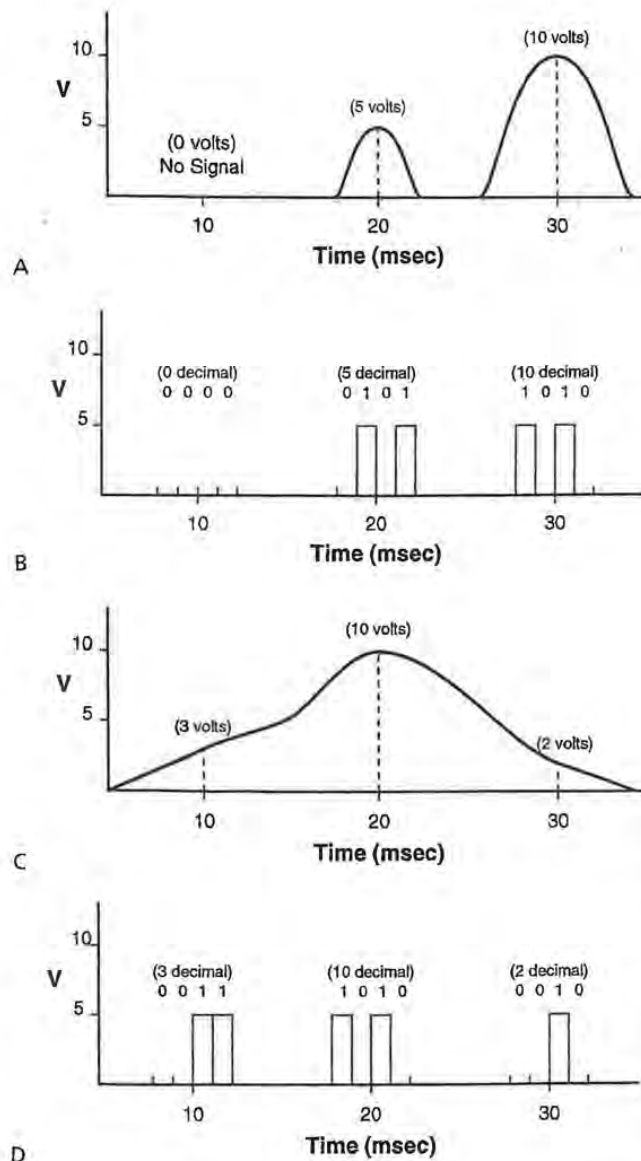


FIGURE 4-2. Analog and digital representation of numerical data. **A:** Three analog voltage pulses, similar to those produced by a photomultiplier tube attached to a scintillator. The height of each pulse represents a number. **B:** These same numbers represented in digital form. **C:** A continuously varying analog signal, such as that from the video camera in a fluoroscopy system. The height of the signal at each point in time represents a number. **D:** The values of this signal, sampled at three points, represented in digital form.

when signals are amplified, attenuation losses, and electronic noise—small stray voltages that exist on circuits and become superimposed on the data. The more the data are transferred, the more distorted they become. On the other hand, data stored or transferred in digital form are remarkably immune to the accumulation of error because of signal distortion. These distortions are seldom of sufficient amplitude to cause a 0 to be mistaken for a 1 or vice versa. Furthermore, most digital circuits do not amplify the incoming data, but make a fresh copy of it, thus preventing distortions from accumulating during multiple transfers.

The digital form facilitates other safeguards. When information integrity is critical, additional redundant information can be sent with each group of bits to permit the receiving device to detect errors or even correct them. A simple error detection method uses parity bits. An additional bit is transmitted with each group of bits, typically with each byte. The bit value designates whether an even or an odd number of bits were in the "1" state. The receiving device determines the parity and compares it with the received parity bit. If the parity of the data and the parity bit do not match, an odd number of bits in the group have errors.

Data can often be transferred in less time using the analog form. However, digital circuits are likely to be less expensive.

Conversion of Analog Data to Digital Form

The transducers, sensors, or detectors of most electronic measuring equipment, including medical imaging devices, produce analog data. If such data are to be analyzed by a computer or other digital equipment, they must be converted into digital form. Devices that perform this function are called *analog-to-digital converters* (ADCs). ADCs are essential components of multichannel analyzers, modern nuclear medicine scintillation cameras, computed radiography systems, modern ultrasound systems, MRI scanners, and CT scanners.

Most analog signals are continuous in time, meaning that at every point in time the signal has a value. However, it is not possible to convert the analog signal to a digital signal at every point in time. Instead, certain points in time must be selected at which the conversion is to be performed. This process is called *sampling*. Each analog sample is then converted into a digital signal. This conversion is called *digitization* or *quantization*.

An ADC is characterized by its *sampling rate* and the *number of bits of output* it provides. The sampling rate is the number of times a second that it can sample and digitize the input signal. Most radiologic applications require very high sampling rates. An ADC produces a digital signal of a fixed number of bits. For example, one may purchase an 8-bit ADC, a 10-bit ADC, or a 12-bit ADC. The number of bits of output is just the number of bits in the digital number produced each time the ADC samples and quantizes the input analog signal.

The digital representation of data is superior to analog representation in its resistance to the accumulation of errors. However, there are also disadvantages to digital representation, an important one being that *the conversion of an analog signal to digital form causes a loss of information*. This loss is due to both sampling and quantization. Because an ADC samples the input signal, the values of the analog signal between the moments of sampling are lost. If the sampling rate of the ADC is sufficiently rapid that the analog signal being digitized varies only slightly during the intervals between sampling, the sampling error will be small. There is a minimum sampling rate requirement, the *Nyquist limit* (discussed in Chapter 10) that ensures the accurate representation of a signal.

Quantization also causes a loss of information. An analog signal is continuous in magnitude, meaning that it may have any value between a minimum and maximum. For example, an analog voltage signal may be 1.0, 2.5, or 1.7893 V. In contrast, a digital signal is limited to a finite number of possible values, determined by the number of bits used for the signal. As was shown earlier in this chapter, a 1-bit digital signal is limited to two values, a 2-bit signal is limited to four values, and an

TABLE 4-5. MAXIMAL ERRORS WHEN DIFFERENT NUMBERS OF BITS ARE USED TO APPROXIMATE AN ANALOG SIGNAL

Number of Bits	Number of Values	Maximal Quantization Error (%)
1	2	25
2	4	12.5
3	8	6.2
8	256	0.20
12	4,096	0.012

N -bit signal is restricted to 2^N possible values. The quantization error is similar to the error introduced when a number is "rounded off." Table 4-5 lists the maximal percent errors associated with digital signals of different numbers of bits.

There are additional sources of error in analog-to-digital conversion other than the sampling and quantization effects described above. For example, some averaging of the analog signal occurs at the time of sampling, and there are inaccuracies in the quantization process. In summary, a digital signal can only approximate the value of an analog signal, causing a loss of information during analog-to-digital conversion.

No analog signal is a perfect representation of the quantity being measured. Statistical effects in the measurement process and stray electronic voltages ("noise") cause every analog signal to have some uncertainty associated with it. To convert an analog signal to digital form without a significant loss of information content, the ADC must sample at a sufficiently high rate and provide a sufficient number of bits so that the error is less than the uncertainty in the analog signal being digitized. In other words, an analog signal with a large *signal-to-noise ratio* (SNR) requires an ADC providing a large number of bits to avoid reducing the SNR.

Digital-to-Analog Conversion

It is often necessary to convert a digital signal to analog form. For example, to display digital images from a CT scanner on a video monitor, the image information must be converted from digital form to an analog voltage signal. This conversion is performed by a *digital-to-analog converter* (DAC). It is important to recognize that the information lost by analog-to-digital conversion is not restored by sending the signal through a DAC (Fig. 4-3).

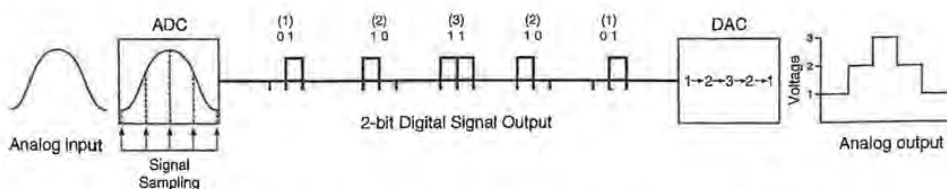


FIGURE 4-3. Analog-to-digital (ADC) conversion and digital-to-analog (DAC) conversion. In this figure, a 2-bit ADC samples the input signal five times. Note that the output signal from the DAC is only an approximation of the input signal to the ADC because the 2-bit digital numbers produced by the ADC can only approximate the continuously varying analog signal.

4.3 COMPONENTS AND OPERATION OF COMPUTERS

A computer consists of a central processing unit (CPU), main memory, and input/output (I/O) devices, all linked together by one or more data pathways called data buses (Fig. 4-4). The main memory of the computer stores the program (sequence of instructions) being executed and the data being processed. The CPU executes the instructions in the program to process the data.

I/O devices enable information to be entered into and retrieved from the computer and to be stored. The I/O devices of a computer usually include a keyboard, a mouse or other pointing device, a video interface and video monitor, several mass storage devices, and often a printer. The keyboard, pointing device, video interface, and video monitor enable the operator to communicate with the computer. Mass storage devices permit the storage of large volumes of programs and data. They include floppy magnetic disk drives, fixed magnetic disk drives, optical disk drives, and magnetic tape drives.

When the CPU or another device sends data to memory or a mass storage device, it is said to be *writing* data, and when it requests data stored in memory or on a storage device, it is said to be *reading* data. Computer memory and data storage devices permit either random access or sequential access to the data. The term *random access* describes a storage medium in which the locations of the data may be read or written in any order, and *sequential access* describes a medium in which data storage locations can only be accessed in a serial manner. Most solid-state memories, magnetic disks, and optical disks are random access, whereas magnetic tape permits only sequential access.

Main Memory

Main memory provides temporary storage for the instructions (computer program) currently being executed and the data currently being processed by the computer.

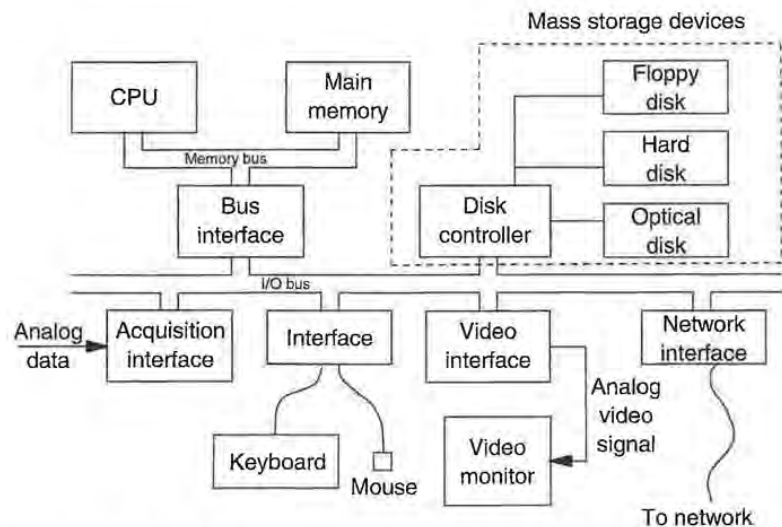


FIGURE 4-4. Components of a computer.

Main memory is used for these functions instead of mass storage devices because the data transfer rate between the CPU and main memory is much faster than that between the CPU and the mass storage devices.

Main memory consists of a large number of data storage locations, each of which contains the same number of bits (usually 1 byte). A unique number, called a *memory address*, identifies each storage location. Memory addresses usually start at zero and increase by one to one less than the total number of memory locations, as shown in Fig. 4-5. A memory address should not be confused with the data stored at that address.

When a CPU performs a memory write, it sends both a memory address, designating the desired storage location, and the data to the memory. When it performs a memory read, it sends the address of the desired data to the memory and, in return, is sent the data at that location.

One type of memory is commonly called *random access memory* (RAM). In common usage, RAM refers to memory onto which the CPU can both read and write data. A better name for such memory is read-write memory. A disadvantage of most read-write memory is that it is *volatile*, meaning that data stored in it are lost when electrical power is turned off.

Another type of memory is called *read-only memory* (ROM). The CPU can only read data from ROM; it cannot write or erase data on ROM. The main advantage to ROM is that data stored in it are not lost when power is lost to the computer. ROM is usually employed to store frequently used programs provided by the manufacturer that perform important functions such as setting up the computer for operation when power is turned on. Although the term *RAM* is commonly used to distinguish read-write memory from ROM, ROM does permit random access.

The size of the main memory can affect the speed at which a program is executed. A very large program or one that processes large volumes of data will usually be executed more quickly on a computer with a large memory. A computer with a small

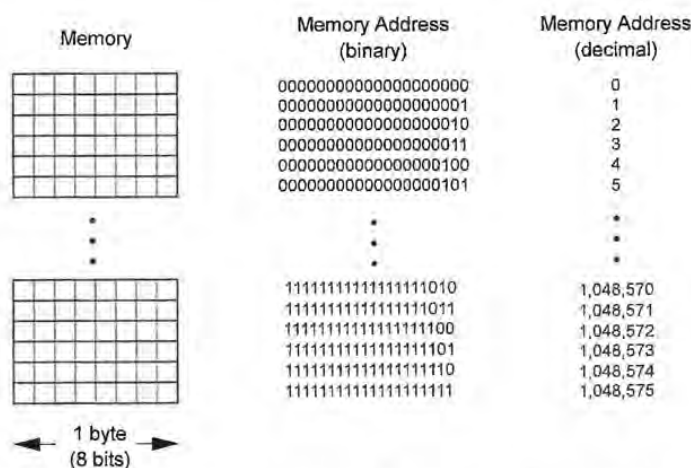


FIGURE 4-5. Main memory, containing 1 megabyte (2^{20} bytes). Each byte in memory is identified by a unique memory address, just as each mailbox in a city is identified by a postal address.

memory can fit only small portions of the program and data into its memory at one time, and must frequently interrupt the processing of data to load other portions of the program and data into the memory from the magnetic disk or other storage devices.

It would be very expensive to have a large main memory that could provide program instructions and data to the CPU as fast as the CPU can process them. The main memories of most computers today use a relatively slow but inexpensive type of memory called dynamic RAM (DRAM). DRAM uses one capacitor (an electrical charge storage device) and one transistor for each bit. It is called dynamic because each bit must be read and refreshed several times a second to maintain the data. To compensate for the low speed of DRAM, many computers contain one or more smaller but faster memories, called *cache memories*, between the main memory and the CPU. They maintain exact copies of small portions of the program and data in the main memory. A cache memory usually consists of static RAM (SRAM), which uses several transistors to represent each bit. SRAM stores data while power is on without requiring refreshing and can be read or written several times faster than DRAM, but requires more space and is more expensive. Cache memories are effective because program segments and data that were recently read from the memory by the CPU are very likely to be read again soon.

Computer Program and Central Processing Unit

A computer program is a sequence of instructions and the central processing unit (CPU) is a set of electronic circuits that executes them. A CPU fetches and executes the instructions in a program sequentially; it fetches an instruction from main memory, executes that instruction, and then fetches the next instruction in the program sequence. A CPU contained in a single computer chip is called a *microprocessor*.

The CPU contains a small number of data storage locations, called *storage registers*, for storing data and memory addresses. (For example, one computer system has 16 registers, each containing 32 bits.) Operations can be performed on data in the registers much more quickly than on data in memory. The CPU also contains an *arithmetic logic unit* (ALU) that performs mathematical operations and comparisons between data stored in the registers or memory. In addition, the CPU contains a clock. This clock is not a time-of-day clock (although the computer may have one of these as well), but a circuit that produces a periodic stream of logic pulses that serve as timing marks to synchronize the operation of the various components of the computer. CPU clock speeds are usually described in millions or billions of hertz (Hz); an Hz is defined as one cycle per second.

An instruction can cause the CPU to perform one or a combination of four actions:

1. Transfer a unit of data (e.g., a 16- or 32-bit word) from a memory address, CPU storage register, or I/O device to another memory address, register, or I/O device.
2. Perform a mathematical operation between two numbers in storage registers in the CPU or in memory.
3. Compare two numbers or other pieces of data in registers or in memory.
4. Change the address of the next instruction in the program to be executed to another location in the program. This type of instruction is called a *branching instruction*.

A CPU fetches and executes the instructions in a program sequentially until it fetches a branching instruction. The CPU then jumps to a new location in the program that is specified by the branching instruction and, starting at the new location, again executes instructions sequentially. There are two types of branching instructions, unconditional and conditional. *Conditional branching instructions* cause a branch to occur only if a specified condition is met. For example, a conditional branching instruction may specify that a branch is to occur only if the number stored in a particular CPU storage register is zero. Conditional branching instructions are especially important because they permit the computer to make "decisions" by allowing the data to alter the order of instruction execution.

Each instruction in a program has two parts. The first part specifies the operation, such as a data transfer, an addition of two numbers, or a conditional branch, to be performed. The second part describes the location of the data to be operated on or compared and the location where the result is to be stored. In the case of an addition, it contains the locations of the numbers to be added and the location where the sum is to be stored. In the case of a branching instruction, the second part is the memory address of the next instruction to be fetched and executed.

The CPU is a bottleneck in a conventional computer. A computer may have many megabytes of data to be processed and a program requiring the execution of instructions many billions of times. However, the CPU executes the instructions only one or a few at a time. One way to increase speed is to build the circuits of the CPU and memory with faster components so the computer can run at a higher clock speed. This makes the computer more expensive and there are practical and technical limits to how fast a circuit can operate.

The other method of relieving the bottleneck is to design the CPU to perform *parallel processing*. This means that the CPU performs some tasks simultaneously rather than sequentially. Most modern CPUs incorporate limited forms of parallel processing. Some specialized CPUs, such as those in array processors (to be described later in this chapter), rely heavily on parallel processing. One form of parallel processing is *pipelining*. Several steps, each requiring one clock cycle, are required for a CPU to execute a single instruction. Older CPUs, such as the Intel 8088, 80286, and 80386 microprocessors, completed only one instruction every few cycles. The instruction execution circuits of newer microprocessors are organized into pipelines, a pipeline having several stages that operate simultaneously. Instructions enter the pipeline one at a time, and are passed from one stage to the next. Several instructions are in different stages of execution at the same time. It takes a several clock cycles to fill the pipeline, but, once full, it completes the execution of one instruction per cycle. Pipelining permits a CPU to attain an average instruction execution rate of almost one instruction per cycle. (Pipelining does not yield exactly one instruction per cycle because some instructions interfere with its operation, in particular ones that modify the data to be operated on by immediately subsequent instructions, and branching instructions, because they change the order of instruction execution.) Superscalar CPU architectures incorporate several pipelines in parallel and can attain average instruction execution rates of two to five instructions per CPU clock cycle.

Input/Output Bus and Expansion Slots

Data buses were described previously in this chapter. The input/output buses of most computers are provided with several expansion slots into which printed circuit

boards can be plugged. The boards installed in these slots can include modems, cards to connect the computer to a computer network, controllers for mass storage devices, display interface cards to provide a video signal to a video monitor, sound cards to provide audio signals to speakers, and acquisition interfaces to other devices such as scintillation cameras. The provision of expansion slots on this bus makes it possible to customize a general-purpose computer for a specific application and to add additional functions and capabilities.

Mass Storage Devices

Mass storage devices permit the nonvolatile (i.e., data are not lost when power is turned off) storage of programs and data. Mass storage devices include floppy disk drives, hard disk drives (nonremovable hard disks are called fixed disks), magnetic tape drives, and optical (laser) disk units. Each of these devices consists of a mechanical drive; the storage medium, which may be removable; and an electronic controller. Despite the wide variety of storage media, there is not one best medium for all purposes, because they differ in storage capacity, data access time, data transfer rate, cost, and other factors.

Magnetic disks are spinning disks coated with a material that may be readily magnetized. Just above the spinning disk is a read/write head that, to read data, senses the magnetization of individual locations on the disk and, to write data, changes the direction of the magnetization of individual locations on the disk. A disk drive has a read/write head on each side of a platter so that both sides can be used for data storage. The data is stored on the disk in concentric rings called *tracks*. The read/write heads move radially over the disk to access data on different tracks. The *access time* of a disk is the time required for the read/write head to reach the proper track (head seek time) and for the spinning of the disk to bring the data of interest to the head (rotational latency). The *data transfer rate* is the rate at which data are read from or written to the disk once the head and disk are in the proper orientation; it is primarily determined by the rotational speed of the disk and the density of data storage in a track.

A typical hard disk drive, as shown in Fig. 4-6, has several rigid platters stacked above each other on a common spindle. The platters continuously rotate at a high speed (typically 5,400 to 15,000 rpm). The read/write heads aerodynamically float microns above the disk surfaces on air currents generated by the spinning platters. A spinning hard disk drive should not be jarred because it might cause a "head crash" in which the head strikes the disk, gouging the disk surface and destroying the head, with a concomitant loss of data. Disks must also be kept clean; a hair or speck of dust can damage the disk and head. For this reason, the portion containing the disks and read/write heads is usually sealed to keep dirt out. Hard disk drives with removable disk platters are less common. The storage capacities of hard disks are very large, ranging from approximately 10 to 180 gigabytes. Although their access times and data transfer rates are very slow compared to memory, they are much faster than those of floppy disks or magnetic tape. For these reasons, hard disks are used on most computers to store frequently used programs and data.

Floppy disks are removable plastic disks coated with a magnetizable material. They are operationally similar to hard disks, but their rotation is intermittent (spinning only during data access), they spin more slowly, and the read/write heads maintain physical contact with the disk surfaces. Their access times are considerably longer and their data transfer rates are much slower than those of hard disks. Their storage capacities, typi-



FIGURE 4-6. Hard-disk drive. A read/write head is visible at the end of the arm overlying the top disk platter. (Photograph courtesy of Seagate Technology, Inc.)

cally about 1.4 megabytes, are much smaller. They are commonly used for sending programs and data from one site to another and for making "backup" copies of important programs and data. Floppy disk drives with capacities of 100 megabytes or more have been developed. However, there are not yet standards for them, and in general those from one manufacturer cannot be read by disk drives of other manufacturers.

Magnetic tape is plastic tape coated with a magnetizable substance. Its average data access times are very long, because the tape must be read serially from the beginning to locate a particular item of data. There are several competing tape-cartridge formats available today. A single tape cartridge can store a very large amount of data, in the range of 4 to 110 gigabytes. Common uses are to "back up" (make a copy for the purpose of safety) large amounts of important information and for archival storage of digital images.

An optical disk is a removable disk that rotates during data access and from which data are read and written using a laser. There are three categories of optical disks—read-only; write-once, read-many-times (WORM); and rewritable. Read-only disks are encoded with data that cannot be modified. To read data, the laser illuminates a point on the surface of the spinning disk and a photodetector senses the intensity of the reflected light. CD-ROMs are an example of read-only optical disks. Most CD-ROMs have capacities of 650 megabytes. CD-ROMs are commonly used by vendors to distribute commercial software and music. WORM devices permit any portion of the disk surface to be written upon only once, but to be read many times. This largely limits the use of WORM optical disks to archival storage. To store data, a laser at a high intensity burns small holes in a layer of the disk. To read data, the laser, at a lower intensity, illuminates a point on the spinning disk and a photodetector senses the intensity of the reflected light. CD-Rs (for recordable) are a common type of WORM disk. Rewritable optical disks permit the stored data to be changed many times. There are currently two types of rewritable optical disks—phase-change and magneto-

optical. The recording material of a phase-change disk has the property that, if heated to one temperature and allowed to cool, it becomes crystalline, whereas if heated to another temperature, it cools to an amorphous phase. To write data on a phase-change disk, the laser heats a point on the recording film, changing its phase from crystalline to amorphous or vice versa. The transparency of the amorphous material differs from that in the crystalline phase. Above the layer of recording material is a reflective layer. Data are read as described above for WORM disks. A magneto-optical disk contains a layer of recording material whose magnetization may be changed by a nearby magnet if heated above a temperature called the *Curie temperature*. Writing data on a magneto-optical disk requires a small electromagnet above the disk and a laser below. The laser heats a small domain on the layer of recording material above the Curie temperature, causing its magnetization to align with that of the electromagnet. When the domain cools below the Curie temperature, the magnetism of that domain becomes "frozen." The magneto-optical disk is read by the laser at a lower intensity. The polarization of the reflected light varies with the magnetization of each domain on the disk. CD-RWs are a common type of rewritable optical disk. The CD is likely to be displaced by a new standard for optical disks, the DVD, that provides a much larger storage capacity. DVDs are available in read-only, recordable, and rewritable forms. ("DVD-RAM" is one of two competing formats for rewritable DVDs.) Advantages of optical disks include large storage capacity, on the order of 650 megabytes to several gigabytes, and much faster data access than magnetic tape. Optical disks provide superior data security than most other media because they are not subject to head crashes, as are magnetic disks; are not as vulnerable to wear and damage as magnetic tape; and are not vulnerable to magnetic fields.

TABLE 4-6. COMPARISON OF CHARACTERISTICS OF MASS STORAGE DEVICES AND MEMORY^a

	Removable	Storage Capacity	Access Time (Average)	Transfer Rate	Cost per Disk/Tape	Media Cost per GB
Floppy disk	Yes	1.2 to 1.4 MB	300 msec	0.03 MB/sec	\$0.30	\$210
Hard disk	Usually not	10 to 182 GB	6 to 15 msec	3 to 50 MB/sec	NA	
Optical disk, CD-ROM, CD-R, CD-RW	Yes	650–700 MB	100 to 150 msec (for 24x)	3.6 MB/sec (for 24x)	\$0.50 (CD-R) \$1.10 (CD-RW)	\$0.80 (CD-R) \$1.70 (CD-RW)
Optical disk, DVD-ROM, DVD-R, DVD-RAM	Yes	4.7 to 17 GB 3.9 GB 5.2 GB		8.1 MB/s	\$24 (DVD-RAM) \$150 (110 GB super DLT)	\$4.60 (DVD-RAM) \$1.35 (110 GB super DLT)
Magnetic tape (cartridge)	Yes	45 MB to 110 GB	Seconds to minutes	0.125 to 10 MB/sec	NA	
Solid-state memory	No	64 MB to 1.5 GB	1 to 80 msec	24 to 250 MB/sec	NA	

^aValues are typical for 2001-vintage computers. Cost refers to one diskette or tape cartridge. Storage capacities and transfer rates are for uncompressed data; they would be higher for compressed data. msec, milliseconds; nsec, nanoseconds (10^6 nsec = 1 msec); MB, megabytes; GB, gigabytes; DLT, digital linear tape.

Table 4-6 compares the characteristics of mass storage devices and memory. Because there is no best device for all purposes, most computer systems today have at least a hard disk drive, a floppy disk drive, and a drive capable of reading optical disks.

Display Interface

An important use of a computer is to display information in visual form. The information may be displayed on a video monitor using a cathode ray tube (CRT) or a flat-panel display. The information may also be presented in hard-copy form, most commonly using photographic film or paper.

Medical imaging computers have video interfaces and display devices, both to assist the user in controlling the computer and to display information. The *video interface* is a device that receives digital data from the memory of a computer via the computer's bus and, using one or more digital-to-analog converters, converts it into an analog video signal for display on a video monitor. Most video interfaces contain their own memory for storing displayed images, although some computers use a portion of main memory for this purpose. The video interface and display devices will be described in greater detail below (see Display).

Keyboard and Pointing Device

Usually, a computer is equipped with keyboard and a pointing device, such as a mouse, trackball, or joystick, to enable the user to direct its operations and to enter data into it.

Acquisition Interface

Until recently, many medical imaging devices, such as scintillation cameras, sent analog data directly to attached computers. In most cases today, the analog data are converted into digital form by the imaging device, which sends the digital data to the computer. A computer's acquisition interface accepts data from an attached device and transfers it to memory via the computer's bus. An acquisition interface that accepts analog data first converts it into digital form using one or more ADCs. Some acquisition interfaces store the received information in their own dedicated memories to permit the computer to perform other tasks during data acquisition. An imaging computer that is not directly connected to an imaging device may instead receive images over a computer network, as described below.

Communications Interface

Computers serve not only as data processing and display devices, but also as communications devices. Thus many, if not most, computers are connected, either intermittently or continuously, to other computers or a computer network. A modem (contraction of "modulator-demodulator") is a device to permit a computer to transfer information to another computer or to a network by analog telephone lines. The modem converts the digital signals from a computer to an analog signal and the modem on the other end converts the analog signal back to digital form. The analog form used by modems is different from the analog signals discussed previously in this chapter, in which the information was encoded as the amplitude of the signal. Instead, modems produce a sinusoidal signal of constant amplitude and encode the information by varying the frequency of the signal (frequency modulation).

A computer may also be directly connected to a network using a network interface card. The card converts digital signals from the computer into the digital form

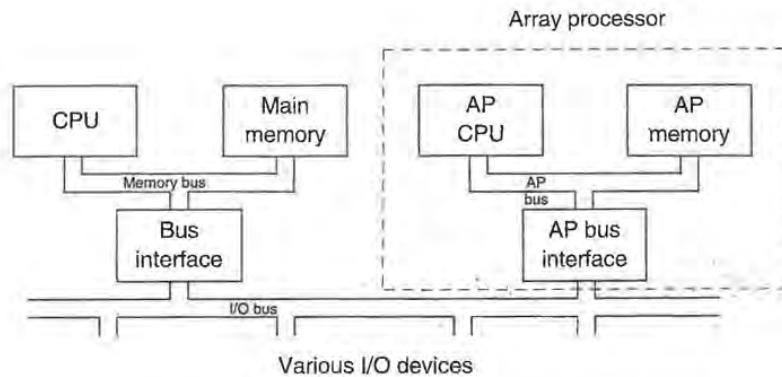


FIGURE 4-7. Computer with an array processor.

required by the network and vice versa. The transfer of information using computer networks is discussed in detail in Chapter 17.

Array Processor

The speed of a general-purpose computer in performing mathematical operations, especially on floating-point numbers, can be greatly enhanced by adding an *array processor*. An array processor is a specialized computer designed to perform the mathematical manipulation of data very quickly. It consists of a CPU, a small high-speed memory, a high-speed bus, and an interface to connect it to the bus of the main computer. The main computer is called the *host computer*. An array processor's CPU relies heavily on parallel processing for its speed. Parallel processing is particularly efficient for processing images. The host computer sends the data to be processed from its memory to the array processor's memory along with the instructions to execute in processing the data. The array processor processes the data and returns the processed data to the memory of the host computer. Figure 4-7 is a diagram of a computer with an array processor.

4.4 PERFORMANCE OF COMPUTER SYSTEMS

There are many factors that affect the time required for a computer to complete a task. These include the speed of the CPU (as reflected its clock speed), how extensively the CPU uses parallel processing, width and the clock speed of the data bus between the CPU and main memory, the size of the memory, and the access times and data transfer rates of mass storage devices. A larger memory permits more of the program being executed and the data being processed to reside in memory, thereby reducing the time that the CPU is idle while awaiting the transfer of program segments and data from mass storage devices into memory. The widths of the data buses affect the speed of the computer; a 32-bit bus can transfer twice as much data per clock cycle as a 16-bit bus.

There are several ways to compare the performance of computer systems. One indicator of CPU speed is the speed of its clock. If two computers are oth-

erwise identical, the computer with the faster clock speed is the fastest. For example, if the CPUs of two computers are both Intel Pentium microprocessors, the CPU running at 1.2 gigahertz (GHz) is faster than the CPU running at 800 megahertz (MHz).

The clock speed of a CPU is not as useful in comparing computers with different CPUs, because one CPU architecture may execute more instructions per clock cycle than another. To avoid this problem, the speed of a CPU may be measured in *millions of instructions per second* (MIPS). If a computer is used to perform extensive mathematical operations, it may be better to measure its performance in millions or billions of floating-point operations per second (megaflops or gigaflops). Another way to compare the performance of computers is to compare the times it takes them to execute standard programs. Programs used for this purpose are called *benchmarks*.

The ability of a computer to execute a particular program in less time than another computer or to perform a greater number of MIPS or megaflops does not prove that it is faster for all applications. For example, the execution speed of one application may rely primarily on CPU speed and CPU-memory transfer rates, whereas another application's speed of execution may largely depend on the rate of transfer of data between the hard disk and main memory. When evaluating computers for a particular purpose, such as reconstructing tomographic images, it is better to compare their speed in performing the actual tasks expected of them than to compare them solely on the basis of clock speed, MIPS, megaflops, or benchmark performance.

4.5 COMPUTER SOFTWARE

Computer Languages

Machine Language

The actual binary instructions that are executed by the CPU are called *machine language*. There are disadvantages to writing a program in machine language. The programmer must have a very detailed knowledge of the computer's construction to write such a program. Also, machine-language programs are very difficult to read and understand. For these reasons, programs are almost never written in machine language.

High-Level Languages

To make programming easier and to make programs more readable, *high-level languages*, such as Fortran, Basic, Pascal, C, and Java, have been developed. These languages enable computer programs to be written using statements similar to simple English, resulting in programs that are much easier to read than those in machine language. Also, much less knowledge of the particular computer is required to write a program in a high-level language. Using a program written in a high-level language requires a program called a *compiler* or an *interpreter* to convert the statements written in the high-level language into machine-language instructions that the CPU can execute. Figure 4-8 shows a simple program in the language Basic.

```
10 I = 1
20 PRINT I
30 I = I+1
40 IF I<11 THEN GOTO 20
50 END
```

FIGURE 4-8. A simple program in the language Basic. The program prints the integers from 1 to 10 on the video monitor.

Applications Programs

An *applications program* is a program that performs a specific function, such as word processing, accounting, scheduling, or image processing. It may be written in machine language or a high-level language. A good applications program is said to be *user-friendly*, meaning that only a minimal knowledge of the computer is needed to use the program.

Operating System

Even when a computer is not executing an assigned program and seems to be idle, there is a fundamental program, called the *operating system*, running. When the user instructs a computer to run a particular program, the operating system copies the program into memory from a disk, transfers control to it, and regains control of the computer when the program has finished. The operating system also handles most of the complex aspects of input/output operations and controls the sharing of resources on multiuser computers. Common operating systems include Windows, Windows NT, and Linux for IBM-compatible PCs; the Mac OS on Apple computers; and Unix for workstations.

Most operating systems store information on mass storage devices as *files*. A single file may contain a program, one or more images, patient data, or a manuscript. Each mass storage device has a *directory* that lists all files on the device and describes the location of each.

The operating system can be used to perform a number of utility functions. These utility functions include listing the files on a mass storage device, determining how much unused space remains on a mass storage device, deleting files, or making a copy of a file on another device. To perform one of these functions, the user types a command or designates a symbol on the computer screen using a pointing device, and the operating system carries it out.

Computer Security

The goals of computer security are to deny unauthorized persons access to confidential information, such as patient data, and to protect software and data from accidental or deliberate modification or loss. A major danger in using computers is that the operating system, applications programs, and important data are often stored on a single disk; an accident could cause all of them to be lost. The primary threats to data and software on computers are mechanical or electrical mal-

function, such as a disk head crash; human error, such as the accidental deletion of a file or the accidental formatting of a disk (causing the loss of all information on the disk); and malicious damage. To reduce the risk of data loss, important files should be copied ("backed up") onto floppy disks, optical disks, or magnetic tape at regularly scheduled times.

Programs written with malicious intent are a threat to computers. The most prevalent of these is the computer virus. A virus is a string of instructions hidden in a program. If a program containing a virus is loaded into a computer and executed, the virus copies itself into other programs stored on mass storage devices. If a copy of an infected program is sent to another computer and executed, that computer becomes infected. Although a virus need not do harm, some deliberately cause damage, such as the deletion of all files on the disk on Friday the 13th. Viruses that are not intended to cause damage may interfere with the operation of the computer or cause damage because they are poorly written. There are three types of viruses—executable file infectors, boot-sector viruses, and macroviruses. Executable file infectors infect files containing executable instructions. An infected file must be imported and executed to trigger one of these. Boot-sector viruses infect the boot sectors of floppy disks. A boot sector is a small portion of a disk that contains instructions to be executed by the computer when power is first turned on. Turning power on to an IBM-compatible personal computer, if an infected floppy is in the floppy drive, will cause an infection. Macroviruses infect macros, programs attached to some word processing and other files. A computer cannot be infected with a virus by the importation of data alone or by the user reading an e-mail message. However, a computer can become infected if an infected file is attached to an e-mail message and if that file is executed.

There are other types of malicious programs. These include Trojan horses, programs that appear to serve one function, but have a hidden purpose; worms, programs that automatically spread over computer networks; and password grabbers, programs that store the passwords of persons logging onto computers for use by an unauthorized person. The primary way to reduce the chance of a virus infection or other problem from malicious software is to establish a policy forbidding the loading of storage media and software from untrustworthy sources. Commercial virus-protection software, which searches files on storage devices and files received over a network for known viruses and removes them, should be used. The final line of defense, however, is the saving of backup copies. Once a computer is infected, it may be necessary to reformat all disks and reload all software and data from the backup copies.

More sophisticated computer operating systems, such as Unix, provide security features including password protection and the ability to grant individual users different levels of access to stored files. Measures should be taken to deny unauthorized users access to your system. Passwords should be used to deny access, directly and over a network or a modem. Computers in nonsecure areas should be "logged off" when not in use. Each user should be granted only the privileges required to accomplish needed tasks. For example, technologists who acquire and process studies and interpreting physicians should not be granted the ability to delete or modify system software files or patient studies, whereas the system manager must be granted full privileges to all files on the system.

4.6 STORAGE, PROCESSING, AND DISPLAY OF DIGITAL IMAGES

Storage

A digital image is a rectangular, usually square, array of numbers. The portion of the image represented by a single number is called a *pixel* (for picture element). For example, in nuclear medicine planar images, a pixel contains the number of counts detected by the corresponding portion of the crystal of a scintillation camera. Figure 4-9 shows

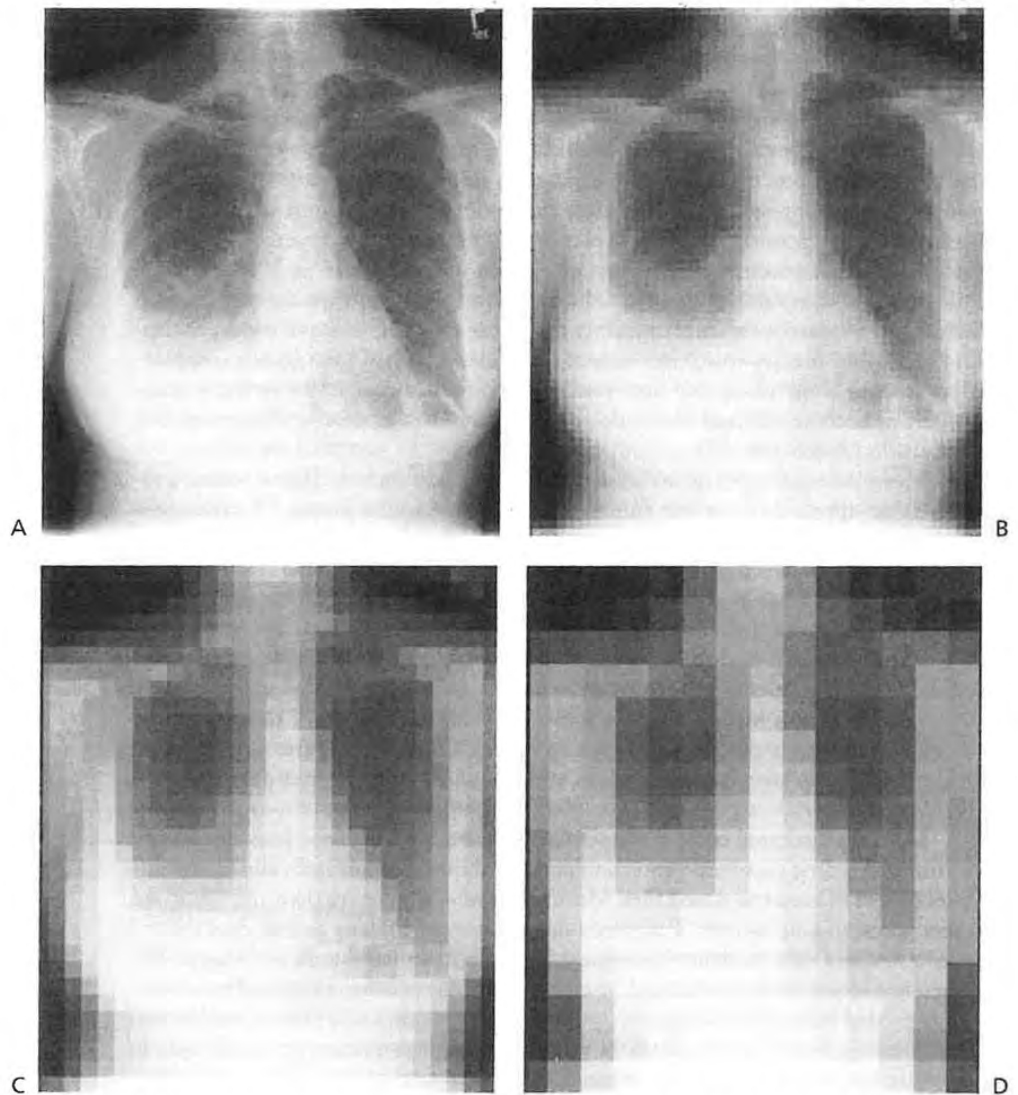


FIGURE 4-9. Effect of pixel size on image quality—digital chest image in formats of 1,024², 64², 32², and 16² pixels (A, B, C, and D, respectively).

TABLE 4-7. TYPICAL RADIOLOGIC IMAGE FORMATS

Modality	Pixel Format	Bits per Pixel
Scintillation camera planar	64 ² or 128 ²	8 or 16
SPECT	64 ² or 128 ²	8 or 16
PET	128 ²	16
Digital fluoroscopy, cardiac catheter lab	512 ² or 1024 ²	8 to 12
Computed radiography, digitized chest films	Typically 2,000 × 2,500	10–12
Mammography (18 × 24 cm) or (24 × 30 cm)	Typically 1,800 × 2,300 to 4,800 × 6,000	12–16
X-ray CT	512 ²	12
MRI	64 ² to 1,024 ²	12
Ultrasound	512 ²	8

CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography; SPECT, single photon emission computed tomography.

a digital image in four different pixel formats. The number in each pixel is converted into a visible light intensity when the image is displayed on a video monitor. Typical image formats used in radiology are listed in Table 4-7.

Imaging modalities with higher spatial resolution require more pixels per image so the image format does not degrade the resolution. In general, an image format should be selected so that the pixel size is on the order of the size of the smallest object to be seen. In the case of a fluoroscope with a 23 cm field of view, the 512² pixel format would be adequate for detecting objects as small as about a half a millimeter in size. To detect objects half this size, a larger format, such as 1024 by 1024, should be selected. When it is necessary to depict the shape of an object, such as a microcalcification in x-ray mammography, an image format much larger than that needed to merely detect the object is required. Figure 4-9 shows the degradation of spatial resolution caused by using too small an image format. The penalty for using larger formats is increased storage and processing requirements and slower transfer of images.

The largest number that can be stored in a single pixel is determined by the number of bits or bytes used for each pixel. If 1 byte (8 bits) is used, the maximal number that can be stored in one pixel is 255 ($2^8 - 1$). If 2 bytes (16 bits) are used, the maximal number that can be stored is 65,535 ($2^{16} - 1$). The amount of contrast resolution provided by an imaging modality determines the number of bits required per pixel. Therefore, imaging modalities with higher contrast resolution require more bits per pixel. For example, the limited contrast resolution of ultrasound usually requires only 6 or 7 bits, and so 8 bits are commonly used for each pixel. On the other hand, x-ray CT provides high contrast resolution and 12 bits are required to represent the full range of CT numbers. Figure 4-10 shows the degradation of spatial resolution caused by using too few bits per pixel.

Pixel size is determined by dividing the distance between two points in the subject being imaged by the number of pixels between these two points in the image. It is approximately equal to the field of view of the imaging device divided by the number of pixels across the image. For example, if a fluoroscope has a 23-cm (9-inch) field of view and the images are acquired in a 512-by-512 format, then the approximate size of a pixel is $23 \text{ cm}/512 = 0.45 \text{ mm}$, for objects at the face of the image intensifier.

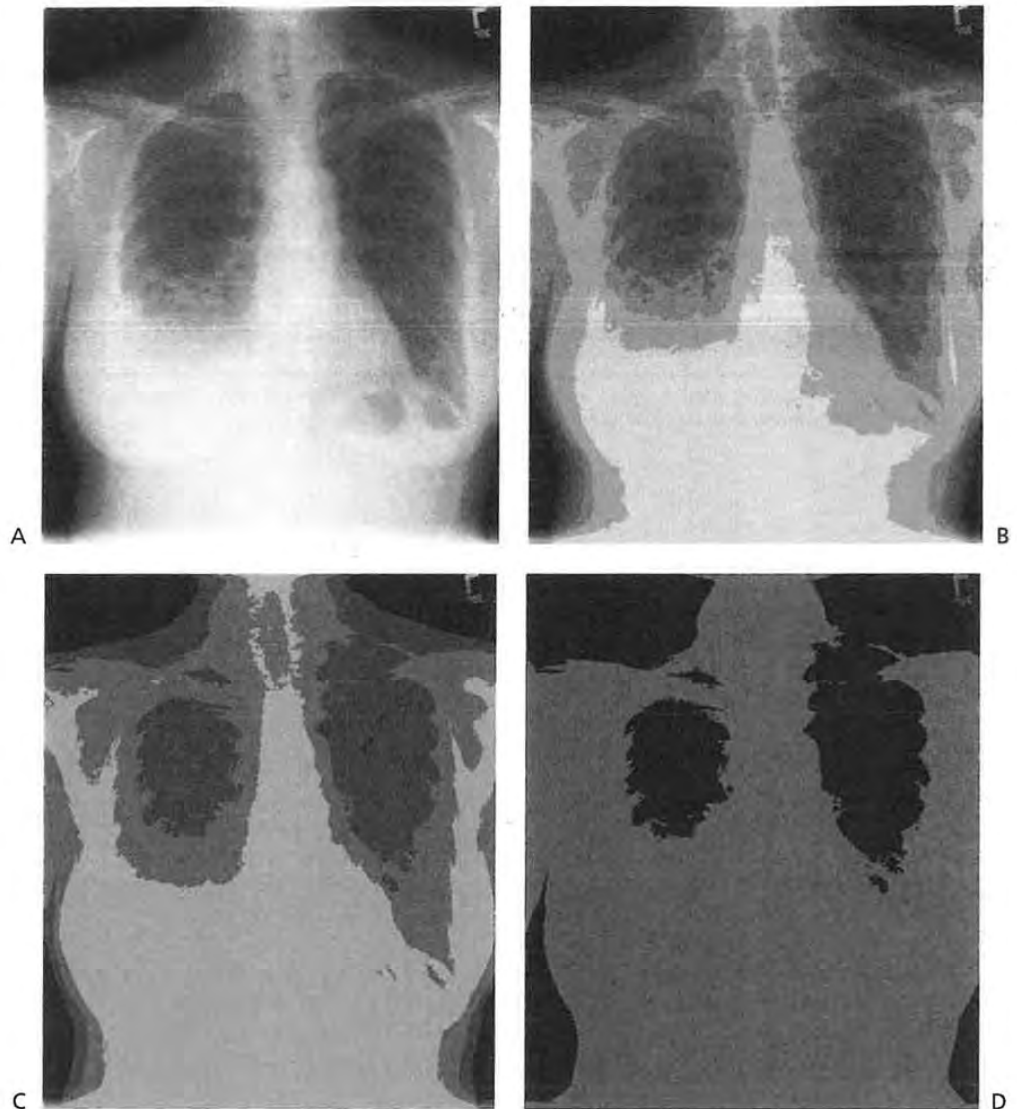


FIGURE 4-10. Effect of number of bits per pixel on image quality—digital chest image in formats of 8, 3, 2, and 1 bits per pixel (**A**, **B**, **C**, and **D**, respectively). Too few bits per pixel not only causes loss of contrast resolution, but also creates the appearance of false contours.

The total number of bytes required to store an image is the number of pixels in the image multiplied by the number of bytes per pixel. For example, the number of kilobytes required to store a 512-by-512 pixel image, if one byte is used per pixel, is

$$(512 \times 512 \text{ pixels}) (1 \text{ byte/pixel}) / (1,024 \text{ bytes/kB}) = 256 \text{ kB.}$$

Similarly, the number of 64-by-64 images that may be stored on a 1.4-megabyte floppy disk, if 16 bits are used per pixel, is

$$(1.4 \text{ MB/disk})(1,024^2 \text{ bytes/MB})/[(64 \times 64 \text{ pixels/image})(2 \text{ bytes/pixel})] = 179 \text{ images/disk}$$

(if no other information is stored on the disk):

Processing

An important use of computers in medical imaging is to process digital images to display the information contained in the images in more useful forms. Digital image processing cannot add information to images. For example, if an imaging modality cannot resolve objects less than a particular size, image processing cannot cause such objects to be seen.

In some cases, the information of interest is visible in the unprocessed images and is merely made more conspicuous. In other cases, information of interest may not be visible at all in the unprocessed image data. X-ray computed tomography is an example of the latter case; observation of the raw projection data fails to reveal many of the structures visible in the processed cross-sectional images. The following examples of image processing are described only superficially and will be discussed in detail in later chapters.

The addition or subtraction of digital images is used in several imaging modalities. Both of these operations require the images being added or subtracted to be of the same format (same number of pixels along each axis of the images) and produce a resultant image (sum image or difference image) in that format. In image subtraction, each pixel in one image is subtracted from the corresponding image in a second image to yield the corresponding pixel value in the difference image. Image addition is similar, but with pixel-by-pixel addition instead of subtraction. Image subtraction is used in digital subtraction angiography to remove the effects of anatomic structures not of interest from the images of contrast-enhanced blood vessels (see Chapter 12) and in nuclear gated cardiac blood pool imaging to yield difference images depicting ventricular stroke-volume and ventricular wall dyskinesia (Chapter 21).

Spatial filtering is used in many types of medical imaging. Medical images often have a grainy appearance, called *quantum mottle*, caused by the statistical nature of the acquisition process. The visibility of quantum mottle can be reduced by a spatial filtering operation called *smoothing*. In most spatial smoothing algorithms, each pixel value in the smoothed image is obtained by a weighted averaging of the corresponding pixel in the unprocessed image with its neighbors. Although smoothing reduces quantum mottle, it also blurs the image. Images must not be smoothed to the extent that clinically significant detail is lost. Spatial filtering can also enhance the edges of structures in an image. Edge-enhancement increases the statistical noise in the image. Spatial filtering is discussed in detail in Chapter 11.

A computer can calculate physiologic performance indices from image data. For example, the estimation of the left ventricular ejection fraction from nuclear gated cardiac blood pool image sequences is described in Chapter 21. In some cases, these data may be displayed graphically, such as the time-activity curves of a bolus of a radiopharmaceutical passing through the kidneys.

In x-ray and nuclear CT, sets of projection images are acquired from different angles about the long axes of patients. From these sets of projection images, computers calculate cross-sectional images depicting tissue linear attenuation coefficients (x-ray CT) or radionuclide concentration (SPECT and PET). The methods

by which volumetric data sets are reconstructed from projection images are discussed in detail in Chapters 13 and 22.

From the volumetric data sets produced by tomography, pseudo-three-dimensional images can be obtained, by methods called surface and volume rendering. These are discussed further in Chapter 13.

In *image co-registration*, image data from one medical imaging modality is superimposed on a spatially congruent image from another modality. Image co-registration is especially useful when one of the modalities, such as SPECT or PET, provides physiologic information lacking anatomic detail and the other modality, often MRI or x-ray CT, provides inferior physiologic information, but superb depiction of anatomy. When the images are obtained using two separate devices, co-registration poses formidable difficulties in matching the position, orientation, and magnification of one image to the other. To greatly simplify co-registration, PET and SPECT devices have been developed that incorporate x-ray CT machines.

Computer-Aided Detection

Computer-aided detection, also known as computer-aided diagnosis, uses a computer program to detect features likely to be of clinical significance in images. Its purpose is not to replace the interpreting physician, but to assist by calling attention to structures that might have been overlooked. For example, software is available to automatically locate structures suggestive of masses, clusters of microcalcifications, and architectural distortions in mammographic images. Computer-aided detection can improve the sensitivity of the interpretation, but also may reduce the specificity. (Sensitivity and specificity are defined in Chapter 10, section 10.7.)

Display

Conversion of a Digital Image into an Analog Video Signal

A computer's video interface converts a digital image into an analog video signal that is displayed by a video monitor. The video interface contains one or more digital-to-analog converters (DACs). Most video interfaces also contain memories to store the image or images being displayed. For a digital image to be displayed, it is first sent to this display memory by the CPU. Once it is stored in this memory, the video interface of the computer reads the pixels in the image sequentially, in the raster pattern shown in Fig. 4-11, and sends these digital numbers to the DAC. The

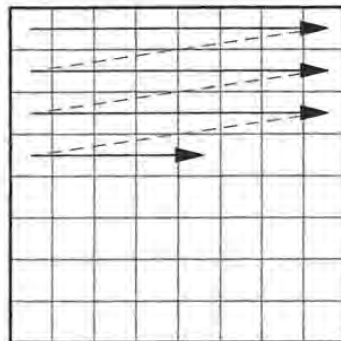


FIGURE 4-11. Order of conversion of pixels in a digital image to analog voltages to form a video signal.

DAC converts each digital number to an analog voltage. Other circuits in the video interface add synchronization pulses to this analog signal, producing an analog video signal that may be displayed on a video monitor. The video interface may contain additional circuits that enable the user to adjust the contrast of the image selected for display.

Gray-Scale Cathode Ray Tube Monitors

Radiographic images, consisting of rectangular arrays of numbers, do not inherently have the property of color and thus the images from most modalities are usually displayed on gray-scale monitors. As will be discussed later, gray-scale monitors provide greater ranges of brightness and thus greater dynamic ranges than do color monitors. However, some modalities, in particular nuclear medicine and ultrasound, use color monitors to provide false color displays (described below). A video monitor may use a cathode ray tube or a flat-panel display to create a visible image.

A cathode ray tube (CRT) is an evacuated container, usually made of glass, with components that generate, focus, and modulate the intensity of one or more electron beams directed onto a fluorescent screen. The CRT of a gray-scale monitor (Fig. 4-12) generates a single electron beam. When a potential difference (voltage) is applied between a pair of electrodes, the negative electrode is called the *cathode* and the positive electrode is called the *anode*. If electrons are released from the cathode, the electric field accelerates them toward the anode. These electrons are called *cathode rays*. The source of the electrons, in most CRTs, is a cathode that is heated by an electrical current, and the fluorescent screen is the anode. A large potential difference, typically 10 to 30 kV, applied between these two electrodes creates an electric field that accelerates the electrons to high kinetic energies and directs them onto the screen. Between the cathode and the screen is a grid electrode. A voltage applied to the grid electrode is used to vary the electric current (electrons per sec-

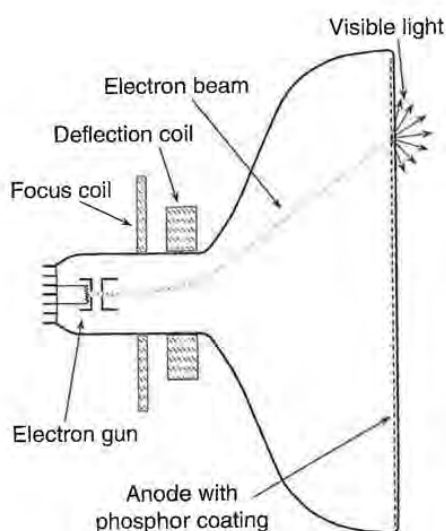


FIGURE 4-12. Gray-scale CRT video monitor. The electron gun contains a heated cathode, the source of the electrons in the beam, and a grid electrode, which modulates the intensity (electrical current) of the beam. This CRT uses an electromagnet to focus the beam and two pairs of electromagnets to steer the beam. (One pair, not shown, is perpendicular to the pair in the diagram.) Alternatively, electrodes placed inside the CRT can perform these functions.

ond) of the beam. The electron beam is focused by either focusing electrodes or an electromagnet into a very narrow cross section where it strikes the screen. The electrons deposit their energy in the phosphor of the screen, causing the emission of visible light. The intensity of the light is proportional to the electric current in the beam, which is determined by the analog video voltage signal applied to the grid electrode. The electron beam can be steered in two perpendicular directions, either by two pairs of electrodes or by two pairs of electromagnets. In a CRT video monitor, the beam is steered in a raster pattern, as shown in Fig. 4-11.

Color CRT Monitors

The CRT of a color monitor is similar to that of a gray-scale monitor, but has three electron guns instead of one, producing three electron beams. The three electron beams are very close to one another and are steered as one, but their intensities are modulated separately. Each pixel of the fluorescent screen contains three distinct regions of phosphors, emitting red, green, and blue light. Just before the screen is a thin metal sheet, called a shadow mask, containing holes positioned so that each electron beam only strikes the phosphor emitting a single color. Thus one of the electron beams creates red light, another creates green light, and the third produces blue light. A mixture of red, green, and blue light, in the proper proportion, can create the perception of any color by the human eye.

Flat-Panel Monitors

Most flat-panel monitors today use liquid crystal displays (LCDs). The liquid crystal material of an LCD does not produce light. Instead, it modulates the intensity of light from another source. The brightest LCDs are backlit. A backlit LCD consists of a uniform light source, typically containing fluorescent tubes and a layer of diffusing material; two polarizing filters; and a layer of liquid crystal material between the two polarizing filters (Fig. 4-13).

Visible light, like any electromagnetic radiation, may be considered to consist of oscillating electrical and magnetic fields, as described in Chapter 2, section 2.1. Unpolarized light consists of light waves whose oscillations are randomly oriented. A polarizing filter is a layer of material that permits components of light waves oscillating in one direction to pass, but absorbs components oscillating in the perpendicular direction. When unpolarized light is incident on a single layer of polarizing material, the intensity of the transmitted light is half that of the incident light. When a second polarizing filter is placed in the beam, parallel to the first filter, the intensity of the beam transmitted through the second filter depends on the orientation of the second filter with respect to the first. If the both filters have the same orientation, the intensity of the beam transmitted through both filters is almost the same as that transmitted through the first. As the second filter is rotated with respect to the first, the intensity of the transmitted light is reduced. When the second filter is oriented so that its polarization is perpendicular to that of the first filter, no light is transmitted.

In an LCD display, the light transmitted through the first filter is polarized. Next, the light passes through a thin layer of liquid crystal material contained between two glass plates. The liquid crystal material consists of rod-shaped molecules aligned so they are parallel. When a voltage is applied across a portion of the

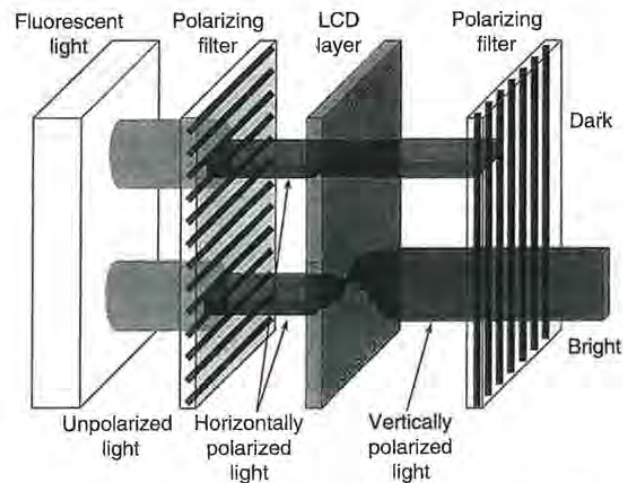


FIGURE 4-13. Two pixels of a gray-scale backlit liquid crystal display video monitor. On each side of the liquid crystal (LC) layer in each pixel is a transparent electrode. In the top pixel, no voltage is applied across the LC layer. The polarization of the light is unchanged as it passes through the LC layer, causing most of the light to be absorbed by the second polarizing filter and thereby producing a dark pixel. In the bottom pixel, a voltage is applied across the LC layer. The applied voltage causes the molecules in the LC layer to twist. The twisted molecules change the polarization of the light, enabling it to pass through the second filter with little reduction in intensity and causing the pixel to appear bright. (The bars shown in the polarizing filters are merely an artist's device to indicate the direction of the polarization, as is depicting the polarized light as a ribbon.)

liquid crystal layer, the molecules twist, changing the polarization of the light by an angle equal to the amount of twist. The light then passes through the second polarizing filter. This filter can be oriented so that pixels are bright when no voltage is applied, or so that pixels are black when no voltage is applied. Figure 4-13 shows the filters oriented so that pixels are dark when no voltage is applied. Thus, when no voltage is applied to a pixel, the light is polarized by the first filter, passes through the liquid crystal material, and then is absorbed by the second polarizing filter, resulting in a dark pixel. As the voltage applied to the pixel is increased, the liquid crystal molecules twist, changing the polarization of the light and decreasing the fraction absorbed by the second filter, thereby making the pixel brighter.

Color LCD monitors have an additional layer containing color filters. Each pixel consists of three adjacent elements, one containing a filter transmitting only red light, the second transmitting green light, and the third transmitting blue light. As mentioned previously, a mixture of red, green, and blue light can create the perception of any color.

In theory, each pixel of a flat-panel display could be controlled by its own electrical pathway. If this were done, a gray-scale display of a thousand by a thousand pixels would contain a million electrical pathways and a color display of the same pixel format would have three times as many. In practice, flat-panel displays are matrix controlled, with one conduction pathway serving each row of pixels and one serving each column. For a thousand by a thousand pixel gray-scale display, only a thousand row pathways and a thousand column pathways are required. A signal is sent to a specific pixel by simultaneously providing voltages to the row conductor and the column conductor for that pixel.

The intensity of each pixel must be maintained while signals are sent to other pixels. In active matrix LCDs, each pixel is provided a transistor and capacitor (or three transistors and capacitors for a color display). The electrical charge stored on the capacitor maintains the voltage signal for the pixel while signals are sent to other pixels. The transistors and capacitors are constructed on a sheet of glass or quartz coated with silicon. This sheet is incorporated as a layer within the LCD. Active matrix LCDs are also called thin film transistor (TFT) displays. TFT technology, without the polarizing filters and liquid crystal material, is used in flat-panel x-ray image receptors and is discussed in Chapter 11. Flat-panel monitors are lighter, require less space, consume less electrical power, and generate less heat than CRT monitors. A disadvantage of LCD displays is that the image can be viewed only from a narrow range of angles, whereas CRT displays can be viewed over a very wide range. Another disadvantage of LCD displays is inoperative pixels. Nearly all LCD monitors have some nonfunctional pixels. These may be always on or always off. The performance characteristics of video monitors are described in Chapter 17.

Contrast Enhancement

To display a digital image, pixel values from the minimal pixel value to the maximal value in the image must be displayed. For displaying the image, a range of video intensities from the darkest to the brightest the video monitor can produce is available. There is a great amount of choice in how the mapping from pixel value to video intensity is to be performed. Selecting a mapping that optimizes the display of important features in the image is called *contrast enhancement*.

There are two methods for contrast enhancement available on most medical image-processing computers: translation table selection and windowing. Both of these may be considered to be methods image processing. However, they are intimately connected with displaying images and so are discussed here. Furthermore, both methods may be applied to the entire image in display memory before the pixel values are sent, one at a time, to form a video signal, or they may be performed on each pixel value, one at a time, after it is selected. The second approach will be described below.

A *translation table* is a lookup table that the video interface uses to modify each pixel value before it is sent to the DAC and becomes an analog signal. Figure 4-14

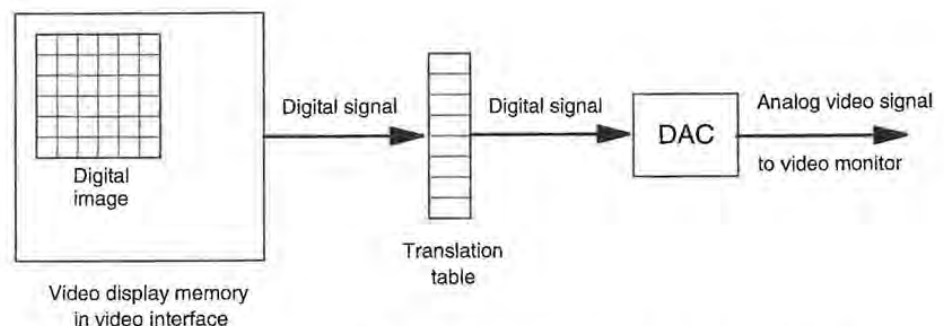


FIGURE 4-14. Video interface showing function of a translation table. An image is transferred to the memory of the video interface for display. The video interface selects pixels in a raster pattern, and sends the pixel values, one at a time, to the lookup table. The lookup table produces a digital value indicating video intensity to the DAC. The DAC converts the video intensity from digital to analog form.

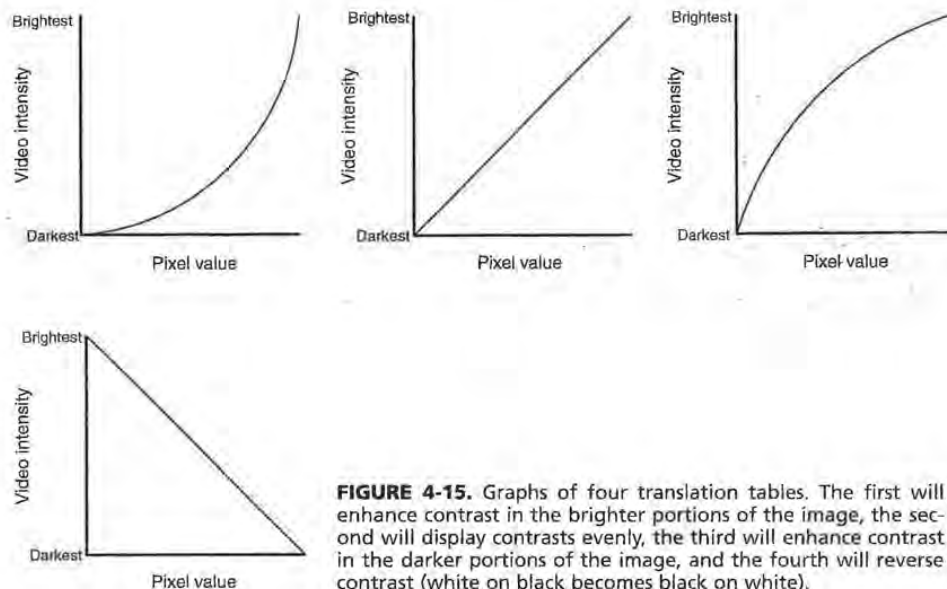


FIGURE 4-15. Graphs of four translation tables. The first will enhance contrast in the brighter portions of the image, the second will display contrasts evenly, the third will enhance contrast in the darker portions of the image, and the fourth will reverse contrast (white on black becomes black on white).

illustrates the use of a translation table. When an image is displayed, each pixel value is sent to the translation table. For every possible pixel value, the table contains a number representing a video intensity. When the translation table receives a pixel value, it looks up the corresponding video intensity and sends this digital number to the DAC. The DAC then converts this number to an analog voltage signal. Figure 4-15 shows four translation tables in graphical form.

Windowing is the second method for contrast enhancement. The viewer may choose to use the entire range of video intensities to display just a portion of the total range of pixel values. For example, an image may have a minimal pixel value of 0 and a maximum of 200. If the user wishes to enhance contrast differences in the brighter portions of the image, a window from 100 to 200 might be selected. Then, pixels with values from 0 to 100 are displayed at the darkest video intensity and so will not be visible on the monitor. Pixels with values of 200 will be displayed at the maximal brightness, and the pixels with values between 100 and 200 will be assigned intensities determined by the current translation table. Figure 4-16 illustrates windowing.

CT and MRI scanners and digital radiography machines usually have level and window controls. The level control determines the midpoint of the pixel values to be displayed and the window control determines the range of pixel values about the level to be displayed. Some nuclear medicine computers require the user to select the lower level (pixel value below which all pixels are displayed at the darkest video intensity) and the upper level (pixel value above which all pixel values are set to the brightest video intensity):

$$\text{Lower level} = \text{Level} - \text{Window}/2$$

$$\text{Upper level} = \text{Level} + \text{Window}/2.$$

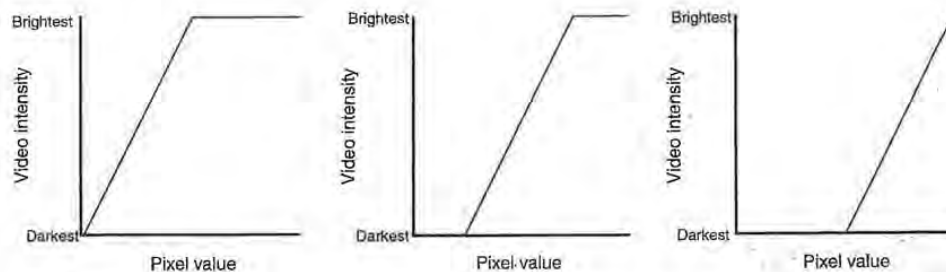


FIGURE 4-16. Windowing an image with pixel values from 0 to 200 to enhance contrast in pixels from 0 to 100 (left), 50 to 150 (center), and 100 to 200 (right).

False Color Displays

As mentioned above, radiographic images do not inherently have the property of color. When color is used to display them, they are called *false color images*. Figure 4-17 shows how false color images are produced. Common uses of false color displays include nuclear medicine, in which color is often used to enhance the perception of contrast, and ultrasound, in which color is used to superimpose flow information on images displaying anatomic information. False color displays are also commonly used in image co-registration, in which an image obtained from one modality is superimposed on a spatially congruent image from another modality.

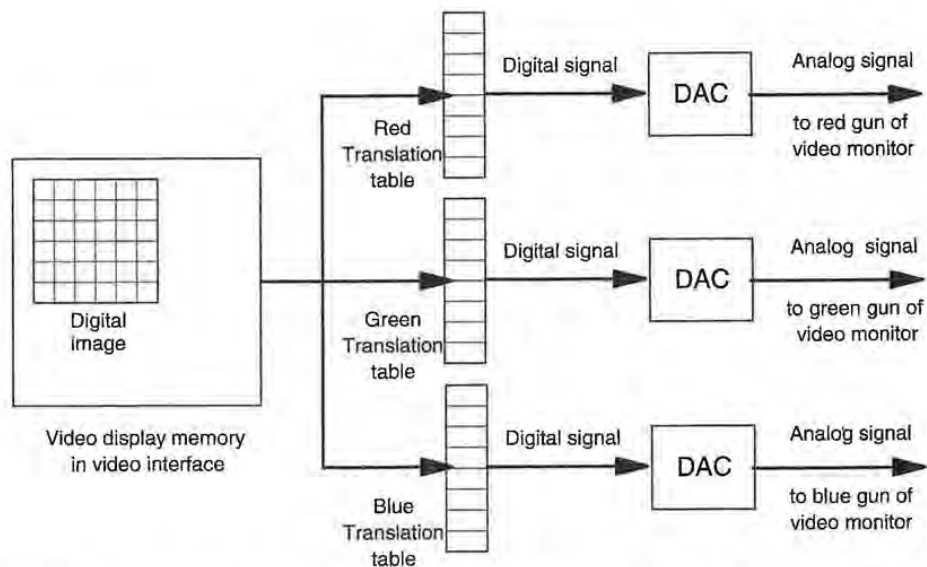


FIGURE 4-17. Creation of a false color display. Each individual pixel value in the image being displayed is used to look up a red, a green, and a blue intensity value. These are simultaneously displayed adjacent to one another within a single pixel of a color video monitor. The mix of these three colors creates the perception of a single color for that pixel.

For example, it is often useful to superimpose nuclear medicine images, which display physiology, but may lack anatomic landmarks, on CT or MRI images. In this case, the nuclear medicine image is usually superimposed in color on a gray-scale MRI or CT image.

Hard-Copy Devices

Digital images may be recorded on photographic film using either a video imager or a laser imager. Most video imagers use the analog video signal produced by the video interface. A video imager consists of a video monitor; a film holder; and one or more lenses, each with a shutter, to focus the video image onto the film, all encased in a light-tight enclosure. A laser imager scans a moving film with a laser beam to produce the image, one row at a time. Most video and laser imagers allow the user to record several images on a single sheet of film. Video and laser imagers are discussed in more detail in Chapter 17.

COMPUTED TOMOGRAPHY

Computed tomography (CT) is in its fourth decade of clinical use and has proved invaluable as a diagnostic tool for many clinical applications, from cancer diagnosis to trauma to osteoporosis screening. CT was the first imaging modality that made it possible to probe the inner depths of the body, slice by slice. Since 1972, when the first head CT scanner was introduced, CT has matured greatly and gained technological sophistication. Concomitant changes have occurred in the quality of CT images. The first CT scanner, an EMI Mark 1, produced images with 80×80 pixel resolution (3-mm pixels), and each pair of slices required approximately 4.5 minutes of scan time and 1.5 minutes of reconstruction time. Because of the long acquisition times required for the early scanners and the constraints of cardiac and respiratory motion, it was originally thought that CT would be practical only for head scans.

CT is one of the many technologies that was made possible by the invention of the computer. The clinical potential of CT became obvious during its early clinical use, and the excitement forever solidified the role of computers in medical imaging. Recent advances in acquisition geometry, detector technology, multiple detector arrays, and x-ray tube design have led to scan times now measured in fractions of a second. Modern computers deliver computational power that allows reconstruction of the image data essentially in real time.

The invention of the CT scanner earned Godfrey Hounsfield of Britain and Allan Cormack of the United States the Nobel Prize for Medicine in 1979. CT scanner technology today is used not only in medicine but in many other industrial applications, such as nondestructive testing and soil core analysis.

13.1 BASIC PRINCIPLES

The mathematical principles of CT were first developed by Radon in 1917. Radon's treatise proved that an image of an unknown object could be produced if one had an infinite number of projections through the object. Although the mathematical details are beyond the scope of this text, we can understand the basic idea behind tomographic imaging with an example taken from radiography.

With plain film imaging, the three-dimensional (3D) anatomy of the patient is reduced to a two-dimensional (2D) projection image. The density at a given point on an image represents the x-ray attenuation properties within the patient along a line between the x-ray focal spot and the point on the detector corresponding to the

point on the image. Consequently, with a conventional radiograph of the patient's anatomy, information with respect to the dimension parallel to the x-ray beam is lost. This limitation can be overcome, at least for obvious structures, by acquiring both a posteroanterior (PA) projection and a lateral projection of the patient. For example, the PA chest image yields information concerning height and width, integrated along the depth of the patient, and the lateral projection provides information about the height and depth of the patient, integrated over the width dimension (Fig. 13-1). For objects that can be identified in both images, such as a pulmonary nodule on PA and lateral chest radiographs, the two films provide valuable location information. For more complex or subtle pathology, however, the two projections are not sufficient. Imagine that instead of just two projections, a series of 360 radiographs were acquired at 1-degree angular intervals around the patient's

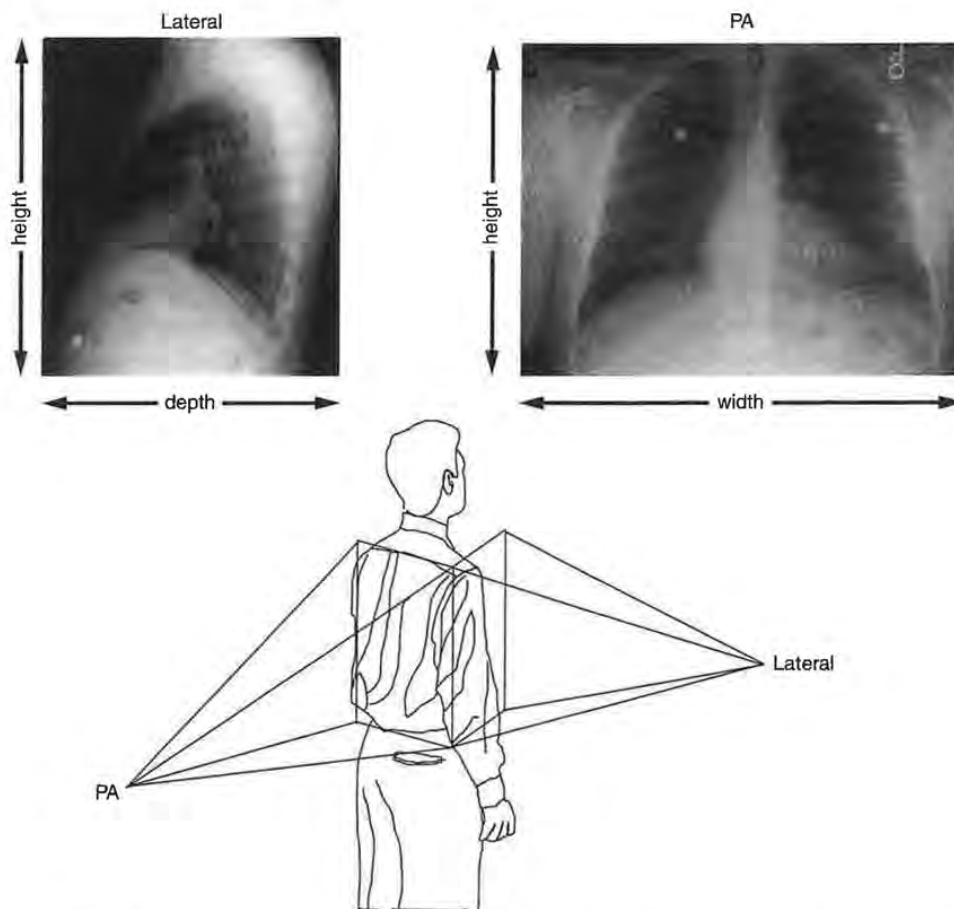


FIGURE 13-1. Posteroanterior and lateral chest radiographs give three-dimensional information concerning the location of an abnormality. In computed tomography, the two views shown here are extended to almost 1,000 views, and with appropriate computer processing true three-dimensional information concerning the patient's anatomy can be displayed.

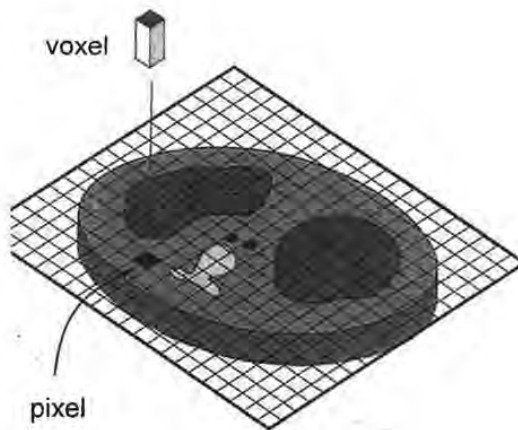


FIGURE 13-2. A pixel (picture element) is the basic two-dimensional element of a digital image. Computed tomographic (CT) images are typically square arrays containing 512×512 pixels, each pixel representing 4,096 possible shades of gray (12 bits). Each pixel in the CT image corresponds to a voxel (volume element) in the patient. The voxel has two dimensions equal to the pixel in the plane of the image, and the third dimension represents the slice thickness of the CT scan.

acic cavity. Such a set of images provides essentially the same data as a thoracic scan. However, the 360 radiographic images display the anatomic information way that would be impossible for a human to visualize: cross-sectional images. These 360 images were stored into a computer, the computer could in principle format the data and generate a complete thoracic CT examination.

The tomographic image is a picture of a slab of the patient's anatomy. The 2D image corresponds to a 3D section of the patient, so that even with CT, three dimensions are compressed into two. However, unlike the case with plain film imaging, the CT slice-thickness is very thin (1 to 10 mm) and is approximately uniform. The 2D array of pixels (short for *picture elements*) in the CT image corresponds to an equal number of 3D voxels (*volume elements*) in the patient. Voxels have the same in-plane dimensions as pixels, but they also include the slice thickness-dimension. Each pixel on the CT image displays the average x-ray attenuation properties of the tissue in the corresponding voxel (Fig. 13-2).

Diagnostic Acquisition

A single transmission measurement through the patient made by a single detector at a given moment in time is called a *ray*. A series of rays that pass through the patient at the same orientation is called a *projection* or *view*. There are two projection geometries that have been used in CT imaging (Fig. 13-3). The more basic type is *parallel beam geometry*, in which all of the rays in a projection are parallel to each other. In *fan beam geometry*, the rays at a given projection angle diverge and have the appearance of a fan. All modern CT scanners incorporate fan beam geometry in the acquisition and reconstruction process. The purpose of the CT scanner is to acquire a large number of transmission measurements through the patient at different positions. The acquisition of a single axial CT image may require approximately 800 rays taken at 1,000 different projection angles, for a total of approximately 800,000 transmission measurements. Before the axial acquisition of the next slice, the table that the patient is lying on is moved slightly in the cranio-caudal direction (the "z-axis" of the scanner), which positions a different slice of tissue in the path of the x-ray beam for the acquisition of the next image.

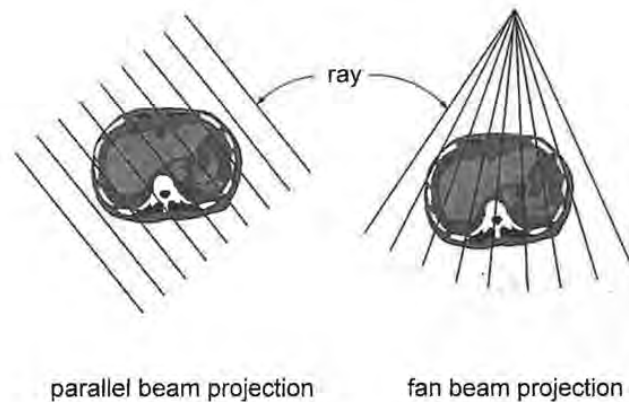


FIGURE 13-3. Computed tomographic (CT) images are produced from a large number of x-ray transmission measurements called rays. A group of rays acquired in a certain geometry is called a *projection* or *view*. Two different geometries have been used in CT, parallel beam projection and fan beam projection, as shown in this figure.

Tomographic Reconstruction

Each ray that is acquired in CT is a transmission measurement through the patient along a line, where the detector measures an x-ray intensity, I_t . The unattenuated intensity of the x-ray beam is also measured during the scan by a reference detector, and this detects an x-ray intensity I_o . The relationship between I_t and I_o is given by the following equation:

$$I_t = I_o e^{-\mu t}$$

where t is the thickness of the patient along the ray and μ is the average linear attenuation coefficient along the ray. Notice that I_t and I_o are machine-dependent values, but the product μt is an important parameter relating to the anatomy of the patient along a given ray. When the equation is rearranged, the measured values I_t and I_o can be used to calculate the parameter of interest:

$$\ln(I_o/I_t) = \mu t$$

where $\ln$ is the natural logarithm (to base e , $e = 2.78 \dots$), t ultimately cancels out, and the value μ for each ray is used in the CT reconstruction algorithm. This computation, which is a preprocessing step performed before image reconstruction, reduces the dependency of the CT image on the machine-dependent parameters, resulting in an image that depends primarily on the patient's anatomic characteristics. This is very much a desirable aspect of imaging in general, and the high clinical utility of CT results, in part, from this feature. By comparison, if a screen-film radiograph is underexposed (I_o is too low) it appears too white, and if it is overexposed (I_o too high) it appears too dark. The density of CT images is independent of I_o , although the noise in the image is affected.

After preprocessing of the raw data, a *CT reconstruction algorithm* is used to produce the CT images. There are numerous reconstruction strategies; however, *filtered backprojection* reconstruction is most widely used in clinical CT scanners. The backprojection method builds up the CT image in the computer by essentially reversing the acquisition steps (Fig. 13-4). During acquisition, attenuation information along a known path of the narrow x-ray beam is integrated by a detector. During backprojection reconstruction, the μ value for each ray is smeared along this same path in the image of the patient. As the data from a large number of rays are

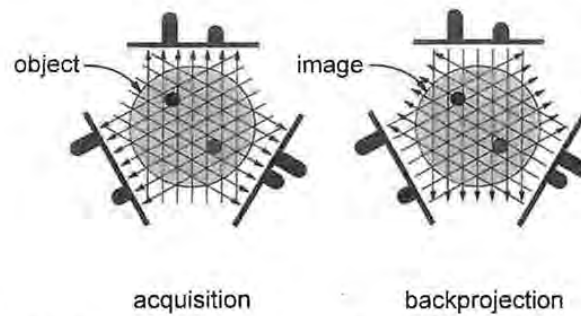


FIGURE 13-4. Data acquisition in computed tomography (CT) involves making transmission measurements through the object at numerous angles around the object (**left**). The process of computing the CT image from the acquisition data essentially reverses the acquisition geometry mathematically (**right**). Each transmission measurement is backprojected onto a digital matrix. After backprojection, areas of high attenuation are positively reinforced through the backprojection process whereas other areas are not, and thus the image is built up from the large collection of rays passing through it.

backprojected onto the image matrix, areas of high attenuation tend to reinforce each other, and areas of low attenuation also reinforce, building up the image in the computer.

13.2 GEOMETRY AND HISTORICAL DEVELOPMENT

Part of understanding an imaging modality involves appreciation of the maturation of that modality. In addition to historical interest, the discussion of the evolution of CT scanners also allows the presentation and reinforcement of several key concepts in CT imaging.

First Generation: Rotate/Translate, Pencil Beam

CT scanners represent a marriage of diverse technologies, including computer hardware, motor control systems, x-ray detectors, sophisticated reconstruction algorithms, and x-ray tube/generator systems. The first generation of CT scanners employed a rotate/translate, pencil beam system (Fig. 13-5). Only two x-ray detectors were used, and they measured the transmission of x-rays through the patient for two different slices. The acquisition of the numerous projections and the multiple rays per projection required that the single detector for each CT slice be physically moved throughout all the necessary positions. This system used parallel ray geometry. Starting at a particular angle, the x-ray tube and detector system translated linearly across the field of view (FOV), acquiring 160 parallel rays across a 24-cm FOV. When the x-ray tube/detector system completed its translation, the whole system was rotated slightly, and then another translation was used to acquire the 160 rays in the next projection. This procedure was repeated until 180 projections were acquired at 1-degree intervals. A total of $180 \times 160 = 28,800$ rays were measured.

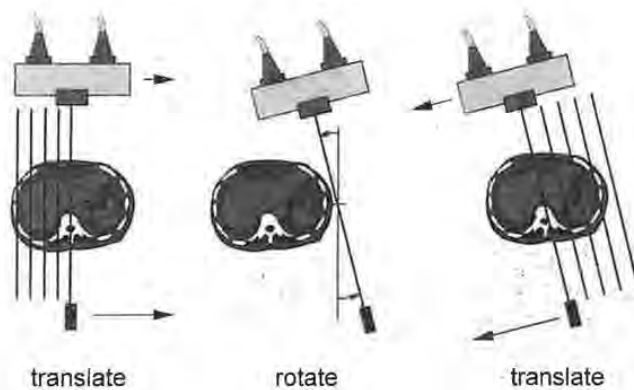


FIGURE 13-5. First-generation (rotate/translate) computed tomography (CT). The x-ray tube and a single detector (per CT slice) translate across the field of view, producing a series of parallel rays. The system then rotates slightly and translates back across the field of view, producing ray measurements at a different angle. This process is repeated at 1-degree intervals over 180 degrees, resulting in the complete CT data set.

As the system translated and measured rays from the thickest part of the head to the area adjacent to the head, a huge change in x-ray flux occurred. The early detector systems could not accommodate this large change in signal, and consequently the patient's head was pressed into a flexible membrane surrounded by a water bath. The water bath acted to bolus the x-rays so that the intensity of the x-ray beam outside the patient's head was similar in intensity to that inside the head. The NaI detector also had a significant amount of "afterglow," meaning that the signal from a measurement taken at one period of time decayed slowly and carried over into the next measurement if the measurements were made temporally too close together.

One advantage of the first-generation CT scanner was that it employed pencil beam geometry—only two detectors measured the transmission of x-rays through the patient. The pencil beam allowed very efficient scatter reduction, because scatter that was deflected away from the pencil ray was not measured by a detector. With regard to scatter rejection, the pencil beam geometry used in first-generation CT scanners was the best.

Second Generation: Rotate/Translate, Narrow Fan Beam

The next incremental improvement to the CT scanner was the incorporation of a linear array of 30 detectors. This increased the utilization of the x-ray beam by 30 times, compared with the single detector used per slice in first-generation systems. A relatively narrow fan angle of 10 degrees was used. In principle, a reduction in scan time of about 30-fold could be expected. However, this reduction time was not realized, because more data ($600 \text{ rays} \times 540 \text{ views} = 324,000 \text{ data points}$) were acquired to improve image quality. The shortest scan time with a second-generation scanner was 18 seconds per slice, 15 times faster than with the first-generation system.

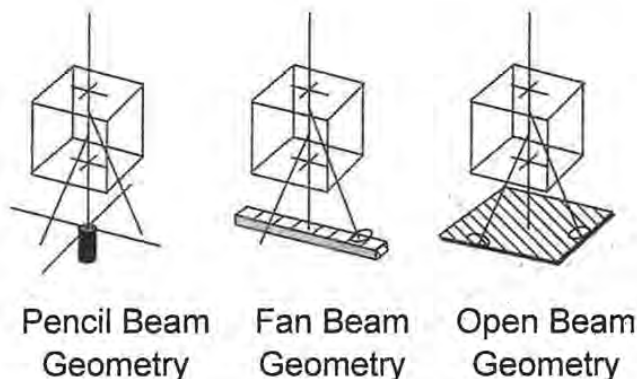


FIGURE 13-6. Pencil beam geometry makes inefficient use of the x-ray source, but it provides excellent x-ray scatter rejection. X-rays that are scattered away from the primary pencil beam do not strike the detector and are not measured. Fan beam geometry makes use of a linear x-ray detector and a divergent fan beam of x-rays. X-rays that are scattered in the same plane as the detector can be detected, but x-rays that are scattered out of plane miss the linear detector array and are not detected. Scattered radiation accounts for approximately 5% of the signal in typical fan beam scanners. Open beam geometry, which is used in projection radiography, results in the highest detection of scatter. Depending on the dimensions and the x-ray energy used, open beam geometries can lead to four detected scatter events for every detected primary photon ($s/p=4$).

Incorporating an array of detectors, instead of just two, required the use of a narrow fan beam of radiation. Although a narrow fan beam provides excellent scatter rejection compared with plain film imaging, it does allow more scattered radiation to be detected than was the case with the pencil beam used in first-generation CT. The difference between pencil beam, fan beam, and open beam geometry in terms of scatter detection is illustrated in Fig. 13-6.

Third Generation: Rotate/Rotate, Wide Fan Beam

The translational motion of first- and second-generation CT scanners was a fundamental impediment to fast scanning. At the end of each translation, the motion of the x-ray tube/detector system had to be stopped, the whole system rotated, and the translational motion restarted. The success of CT as a clinical modality in its infancy gave manufacturers reason to explore more efficient, but more costly, approaches to the scanning geometry.

The number of detectors used in third-generation scanners was increased substantially (to more than 800 detectors), and the angle of the fan beam was increased so that the detector array formed an arc wide enough to allow the x-ray beam to interrogate the entire patient (Fig. 13-7). Because detectors and the associated electronics are expensive, this led to more expensive CT scanners. However, spanning the dimensions of the patient with an entire row of detectors eliminated the need for translational motion. The multiple detectors in the detector array capture the same number of ray measurements in one instant as was required by a complete

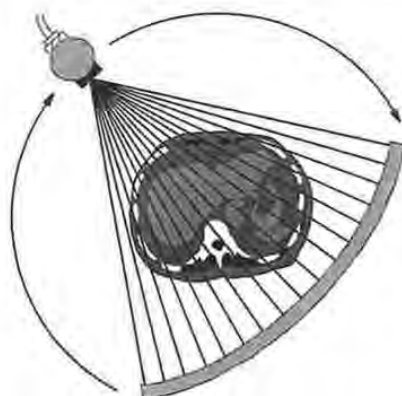


FIGURE 13-7. Third-generation (rotate/rotate) computed tomography. In this geometry, the x-ray tube and detector array are mechanically attached and rotate together inside the gantry. The detector array is long enough so that the fan angle encompasses the entire width of the patient.

translation in the earlier scanner systems. The mechanically joined x-ray tube and detector array rotate together around the patient without translation. The motion of third-generation CT is "rotate/rotate," referring to the rotation of the x-ray tube and the rotation of the detector array. By elimination of the translational motion, the scan time is reduced substantially. The early third-generation scanners could deliver scan times shorter than 5 seconds. Newer systems have scan times of one-half second.

The evolution from first- to second- and second- to third-generation scanners involved radical improvement with each step. Developments of the fourth- and fifth-generation scanners led not only to some improvements but also to some compromises in clinical CT images, compared with third-generation scanners. Indeed, rotate/rotate scanners are still as viable today as they were when they were introduced in 1975. The features of third- and fourth-generation CT should be compared by the reader, because each offers some benefits but also some trade-offs.

Fourth Generation: Rotate/Stationary

Third-generation scanners suffered from the significant problem of ring artifacts, and in the late 1970s fourth-generation scanners were designed specifically to address these artifacts. It is never possible to have a large number of detectors in perfect balance with each other, and this was especially true 25 years ago. Each detector and its associated electronics has a certain amount of drift, causing the signal levels from each detector to shift over time. The rotate/rotate geometry of third-generation scanners leads to a situation in which each detector is responsible for the data corresponding to a ring in the image (Fig. 13-8). Detectors toward the center of the detector array provide data in the reconstructed image in a ring that is small in diameter, and more peripheral detectors contribute to larger diameter rings.

Third-generation CT uses a fan geometry in which the vertex of the fan is the x-ray focal spot and the rays fan out from the x-ray source to each detector on the detector array. The detectors toward the center of the array make the transmission measurement I_1 , while the reference detector that measures I_0 is positioned near the edge of the detector array. If g_1 is the gain of the reference detector, and g_2 is the

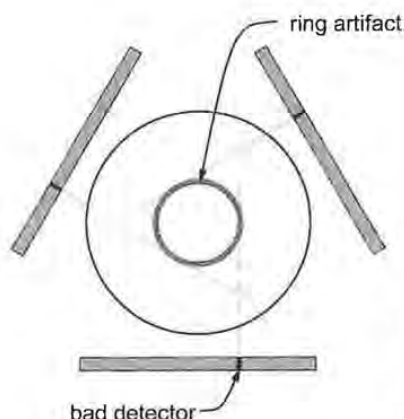


FIGURE 13-8. With third-generation geometry in computed tomography, each individual detector gives rise to an annulus (ring) of image information. When a detector becomes miscalibrated, the tainted data can lead to ring artifacts in the reconstructed image.

gain of the other detector, then the transmission measurement is given by the following equation:

$$\ln(g_1 I_0 / g_2 I_t) = \mu t$$

The equation is true only if the gain terms cancel each other out, and that happens when $g_1 = g_2$. If there is electronic drift in one or both of the detectors, then the gain changes between detectors, so that $g_1 \neq g_2$. So, for third-generation scanners, even a slight imbalance between detectors affects the μt values that are back-projected to produce the CT image, causing the ring artifacts.

Fourth-generation CT scanners were designed to overcome the problem of ring artifacts. With fourth-generation scanners, the detectors are removed from the rotating gantry and are placed in a stationary 360-degree ring around the patient (Fig. 13-9), requiring many more detectors. Modern fourth-generation CT systems use about 4,800 individual detectors. Because the x-ray tube rotates and the detectors are stationary, fourth-generation CT is said to use a rotate/stationary geometry. During acquisition with a fourth-generation scanner, the divergent x-ray beam

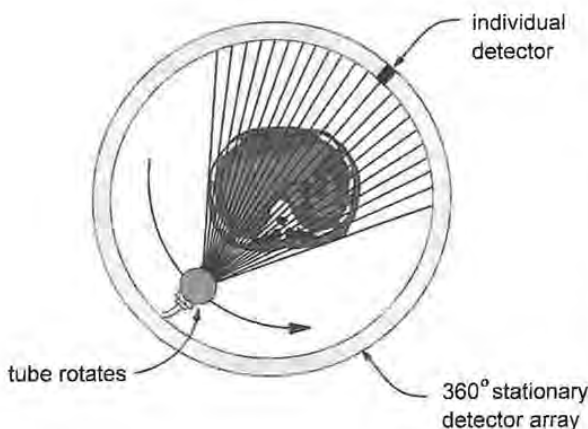


FIGURE 13-9. Fourth-generation (rotate/stationary) computed tomography (CT). The x-ray tube rotates within a complete circular array of detectors, which are stationary. This design requires about six times more individual detectors than a third-generation CT scanner does. At any point during the scan, a divergent fan of x-rays is detected by a group of x-ray detectors.

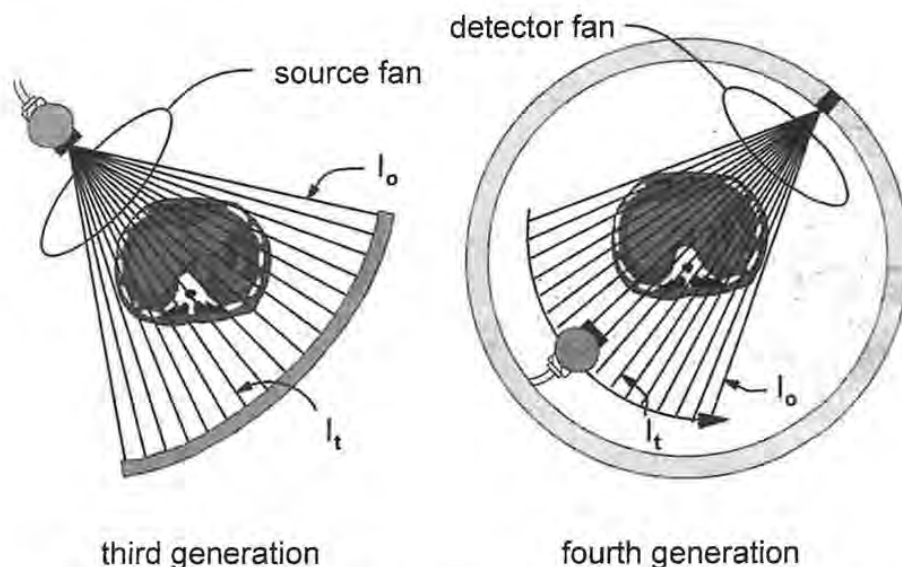


FIGURE 13-10. The fan beam geometry in third-generation computed tomography uses the x-ray tube as the apex of the fan (source fan). Fourth-generation scanners normalize the data acquired during the scan so that the apex of the fan is an individual detector (detector fan). With third-generation scanners, the detectors near the edge of the detector array measure the reference x-ray beam. With fourth-generation scanners, the reference beam is measured by the same detector used for the transmission measurement.

emerging from the x-ray tube forms a fan-shaped x-ray beam. However, the data are processed for fan beam reconstruction with each detector as the vertex of a fan, the rays acquired by each detector being fanned out to different positions of the x-ray source. In the vernacular of CT, third-generation design uses a *source fan*, whereas fourth-generation uses a *detector fan*. The third-generation fan data are acquired by the detector array simultaneously, in one instant of time. The fourth-generation fan beam data are acquired by a single detector over the period of time that is required for the x-ray tube to rotate through the arc angle of the fan. This difference is illustrated in Fig. 13-10. With fourth-generation geometry, each detector acts as its own reference detector. For each detector with its own gain, g , the transmission measurement is calculated as follows:

$$\ln(g I_0 / g I_t) = \mu t$$

Note that the single g term in this equation is guaranteed to cancel out. Therefore, ring artifacts are eliminated in fourth-generation scanners. It should be mentioned, however, that with modern detectors and more sophisticated calibration software, third-generation CT scanners are essentially free of ring artifacts as well.

Fifth Generation: Stationary/Stationary

A novel CT scanner has been developed specifically for cardiac tomographic imaging. This "cine-CT" scanner does not use a conventional x-ray tube; instead, a large arc of tungsten encircles the patient and lies directly opposite to the detector ring.

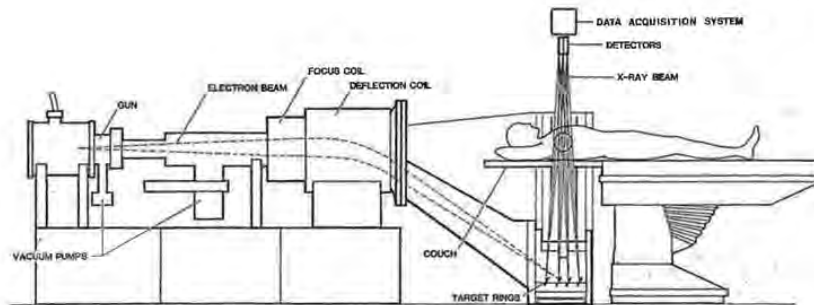


FIGURE 13-11. A schematic diagram of the fifth-generation, Imatron (South San Francisco, CA) cine-CT scanner is shown.

X-rays are produced from the focal track as a high-energy electron beam strikes the tungsten. There are no moving parts to this scanner gantry. The electron beam is produced in a cone-like structure (a vacuum enclosure) behind the gantry and is electronically steered around the patient so that it strikes the annular tungsten target (Fig. 13-11). Cine-CT systems, also called *electron beam scanners*, are marketed primarily to cardiologists. They are capable of 50-msec scan times and can produce fast-frame-rate CT movies of the beating heart.

Sixth Generation: Helical

Third-generation and fourth-generation CT geometries solved the mechanical inertia limitations involved in acquisition of the individual projection data by eliminating the translation motion used in first- and second-generation scanners. However, the gantry had to be stopped after each slice was acquired, because the detectors (in third-generation scanners) and the x-ray tube (in third- and fourth-generation machines) had to be connected by wires to the stationary scanner electronics. The ribbon cable used to connect the third-generation detectors with the electronics had to be carefully rolled out from a cable spool as the gantry rotated, and then as the gantry stopped and began to rotate in the opposite direction the ribbon cable had to be retracted. In the early 1990s, the design of third- and fourth-generation scanners evolved to incorporate *slip ring* technology. A slip ring is a circular contact with sliding brushes that allows the gantry to rotate continually, untethered by wires. The use of slip-ring technology eliminated the inertial limitations at the end of each slice acquisition, and the rotating gantry was free to rotate continuously throughout the entire patient examination. This design made it possible to achieve greater rotational velocities than with systems not using a slip ring, allowing shorter scan times.

Helical CT (also inaccurately called *spiral CT*) scanners acquire data while the table is moving; as a result, the x-ray source moves in a helical pattern around the patient being scanned. Helical CT scanners use either third- or fourth-generation slip-ring designs. By avoiding the time required to translate the patient table, the total scan time required to image the patient can be much shorter (e.g., 30 seconds for the entire abdomen). Consequently, helical scanning allows the use of less contrast agent and increases patient throughput. In some instances the entire scan can be performed within a single breath-hold of the patient, avoiding inconsistent levels of inspiration.

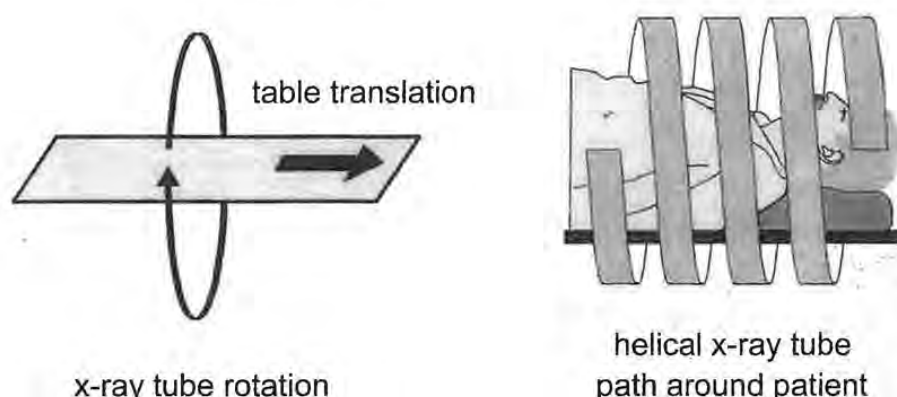


FIGURE 13-12. With helical computed tomographic scanners, the x-ray tube rotates around the patient while the patient and the table are translated through the gantry. The net effect of these two motions results in the x-ray tube traveling in a helical path around the patient.

The advent of helical scanning has introduced many different considerations for data acquisition. In order to produce reconstructions of *planar* sections of the patient, the raw data from the helical data set are interpolated to approximate the acquisition of planar reconstruction data (Fig. 13-12). The speed of the table motion relative to the rotation of the CT gantry is a very important consideration, and the *pitch* is the parameter that describes this relationship (discussed later).

Seventh Generation: Multiple Detector Array

X-ray tubes designed for CT have impressive heat storage and cooling capabilities, although the instantaneous production of x-rays (i.e., x-rays per milliamper-second [mAs]) is constrained by the physics governing x-ray production. An approach to overcoming x-ray tube output limitations is to make better use of the x-rays that are produced by the x-ray tube. When multiple detector arrays are used (Fig. 13-13), the collimator spacing is wider and therefore more of the x-rays that are produced by the x-ray tube are used in producing image data. With conventional, single detector array scanners, opening up the collimator increases the slice thickness, which is good for improving the utilization of the x-ray beam but reduces spatial resolution in the slice thickness dimension. With the introduction of multiple detector arrays, the slice thickness is determined by the detector size and not by the collimator. This represents a major shift in CT technology.

For example, a multiple detector array CT scanner may operate with four contiguous, 5-mm detector arrays and 20-mm collimator spacing. For the same technique (kilovoltage [kV] and mAs), the number of x-rays being detected is four times that of a single detector array with 5-mm collimation. Furthermore, the data set from the 4×5 mm multiple detector array can be used to produce true 5-mm slices, or data from adjacent arrays can be added to produce true 10-, 15-, or even 20-mm slices, all from the same acquisition. The flexibility of CT acquisition protocols and increased efficiency resulting from multiple detector array CT scanners allows better patient imaging; however, the number of parameters involved in the CT acqui-

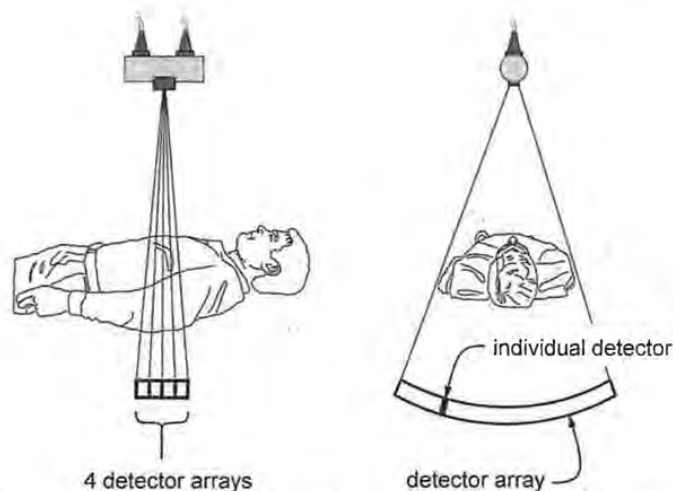


FIGURE 13-13. Multiple detector array computed tomographic (CT) scanners use several, closely spaced, complete detector arrays. With no table translation (nonhelical acquisition), each detector array acquires a separate axial CT image. With helical acquisition on a multiple detector array system, table speed and detector pitch can be increased, increasing the coverage for a given period of time.

sition protocol is increased as well. Also with multiple detector arrays, the notion of helical pitch needs to be redefined. This topic is discussed later in the chapter.

13.3 DETECTORS AND DETECTOR ARRAYS

Xenon Detectors

Xenon detectors use high-pressure (about 25 atm) nonradioactive xenon gas, in long thin cells between two metal plates (Fig. 13-14). Although a gaseous detector does not have the same detection efficiency as a solid one, the detector can be made very thick (e.g., 6 cm) to compensate in part for its relatively low density. The metal

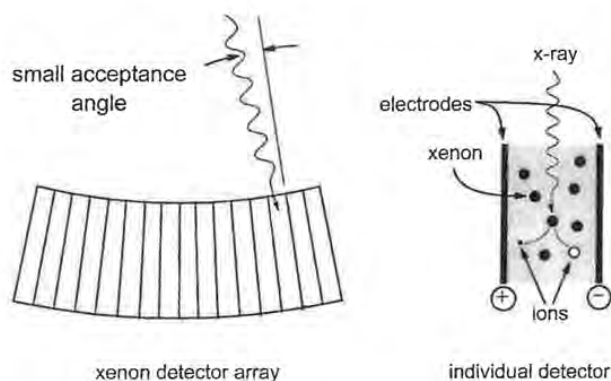


FIGURE 13-14. Xenon detector arrays are a series of highly directional xenon-filled ionization chambers. As x-rays ionize xenon atoms, the charged ions are collected as electric current at the electrodes. The current is proportional to the x-ray fluence.

septa that separate the individual xenon detectors can also be made quite thin, and this improves the geometric efficiency by reducing dead space between detectors. The geometric efficiency is the fraction of primary x-rays exiting the patient that strike active detector elements.

The long, thin ionization plates of a xenon detector are highly directional. For this reason, xenon detectors must be positioned in a fixed orientation with respect to the x-ray source. Therefore, xenon detectors cannot be used for fourth-generation scanners, because those detectors have to record x-rays as the source moves over a very wide angle. Xenon detectors can be used only for third-generation systems.

Xenon detectors for CT are ionization detectors—a gaseous volume is surrounded by two metal electrodes, with a voltage applied across the two electrodes (see Chapter 20). As x-rays interact with the xenon atoms and cause ionization (positive atoms and negative electrons), the electric field (volts per centimeter) between the plates causes the ions to move to the electrodes, where the electronic charge is collected. The electronic signal is amplified and then digitized, and its numerical value is directly proportional to the x-ray intensity striking the detector. Xenon detector technology has been surpassed by solid-state detectors, and its use is now relegated to inexpensive CT scanners.

Solid-State Detectors

A solid-state CT detector is composed of a scintillator coupled tightly to a photodiode. The scintillator emits visible light when it is struck by x-rays, just as in an x-ray intensifying screen. The light emitted by the scintillator reaches the photodiode, typically a photodiode, which is an electronic device that converts light intensity into an electrical signal proportional to the light intensity (Fig. 13-15). This scintillator-photodiode design of solid-state CT detectors is very similar in concept to many digital radiographic x-ray detector systems; however, the performance requirements of CT are slightly different. The detector size in CT is measured in millimeters (typically 1.0×15 mm or 1.0×1.5 mm for multiple detector array scanners), whereas detector elements in digital radiography systems are typi-

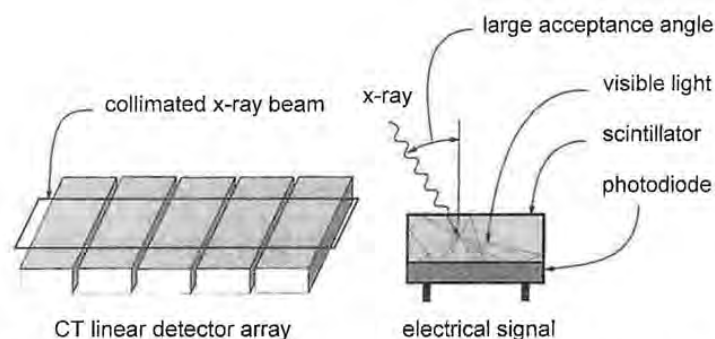


FIGURE 13-15. Solid-state detectors comprise a scintillator coupled to a photodiode. X-rays interact in the scintillator, releasing visible light, which strikes the photodiode and produces an electric signal proportional to the x-ray fluence. For a single CT linear detector array, the detectors are slightly wider than the widest collimated x-ray beam thickness, approximately 12 mm.

cally 0.10 to 0.20 mm on each side. CT requires a very high-fidelity, low-noise signal, typically digitized to 20 or more bits.

The scintillator used in solid-state CT detectors varies among manufacturers, with CdWO_4 , yttrium and gadolinium ceramics, and other materials being used. Because the density and effective atomic number of scintillators are substantially higher than those of pressurized xenon gas, solid-state detectors typically have better x-ray absorption efficiency. However, to reduce crosstalk between adjacent detector elements, a small gap between detector elements is necessary, and this reduces the geometric efficiency somewhat.

The top surface of a solid-state CT detector is essentially flat and therefore is capable of x-ray detection over a wide range of angles, unlike the xenon detector. Solid-state detectors are required for fourth-generation CT scanners, and they are used in most high-tier third-generation scanners as well.

Multiple Detector Arrays

Multiple detector arrays are a set of several linear detector arrays, tightly abutted (Fig. 13-16). The multiple detector array is an assembly of multiple solid-state

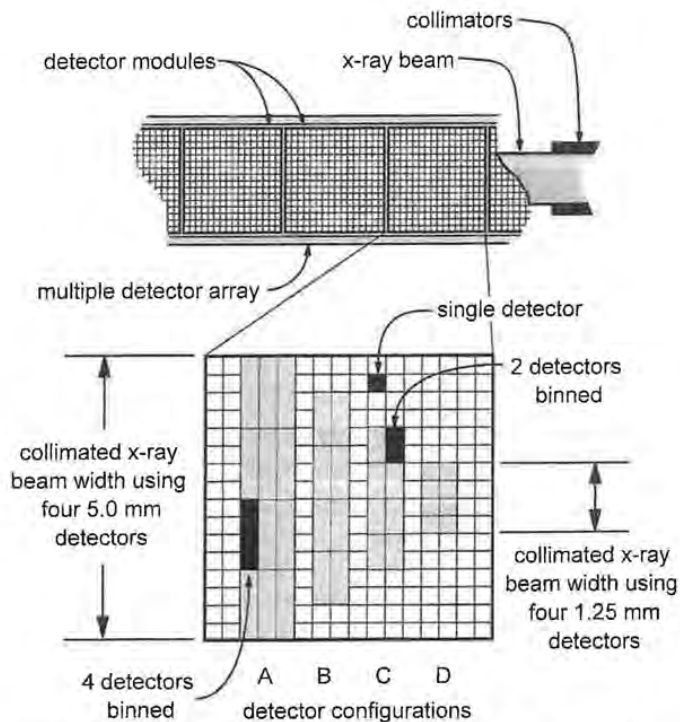


FIGURE 13-16. A multiple detector array composed of multiple detector modules is illustrated in the top panel. In the bottom panel, four different detector configurations are shown, (A-D) each allowing the detection of four simultaneous slices but at different slice widths determined by how many detectors are binned together.

detector array modules. With a traditional single detector array CT system, the detectors are quite wide (e.g., 15 mm) and the adjustable collimator determines slice thickness, typically between 1 and 13 mm. With these systems, the spacing between the collimator blades is adjusted by small motors under computer control. With multiple detector arrays, slice width is determined by the detectors, not by the collimator (although a collimator does limit the beam to the total slice thickness). To allow the slice width to be adjustable, the detector width must be adjustable. It is not feasible, however, to physically change the width of the detector arrays *per se*. Therefore, with multislice systems, the slice width is determined by grouping one or more detector units together. For one manufacturer, the individual detector elements are 1.25 mm wide, and there are 16 contiguous detectors across the module. The detector dimensions are referenced to the scanner's *isocenter*, the point at the center of gantry rotation. The electronics are available for four detector array channels, and one, two, three or four detectors on the detector module can be combined to achieve slices of 4×1.25 mm, 4×2.50 mm, 4×3.75 mm, or 4×5.00 mm. To combine the signal from several detectors, the detectors are essentially wired together using computer-controlled switches. Other manufacturers use the same general approach but with different detector spacings. For example, one manufacturer uses 1-mm detectors everywhere except in the center, where four 0.5-mm-wide detectors are used. Other manufacturers use a gradually increasing spacing, with detector widths of 1.0, 1.5, 2.5, and 5.0 mm going away from the center. Increasing the number of active detector arrays beyond four (used in the example discussed) is a certainty.

Multiple detector array CT scanners make use of solid-state detectors. For a third-generation multiple detector array with 16 detectors in the slice thickness dimension and 750 detectors along each array, 12,000 individual detector elements are used. The fan angle commonly used in third-generation CT scanners is about 60 degrees, so fourth-generation scanners (which have detectors placed around 360 degrees) require roughly six times as many detectors as third-generation systems. Consequently, all currently planned multiple detector array scanners make use of third-generation geometry.

13.4 DETAILS OF ACQUISITION

Slice Thickness: Single Detector Array Scanners

The slice thickness in single detector array CT systems is determined by the physical collimation of the incident x-ray beam with two lead jaws. As the gap between the two lead jaws widens, the slice thickness increases. The width of the detectors in the single detector array places an upper limit on slice thickness. Opening the collimation beyond this point would do nothing to increase slice thickness, but would increase both the dose to the patient and the amount of scattered radiation.

There are important tradeoffs with respect to slice thickness. For scans performed at the same kV and mAs, the number of detected x-ray photons increases *linearly* with slice thickness. For example, going from a 1-mm to a 3-mm slice thickness *triples* the number of detected x-ray photons, and the signal-to-noise ratio (SNR) increases by 73%, since $\sqrt{3} = 1.73$. Increasing the slice thickness from

5 to 10 mm with the same x-ray technique (kV and mAs) *doubles* the number of detected x-ray photons, and the SNR increases by 41%: $\sqrt{2} = 1.41$. Larger slice thicknesses yield better contrast resolution (higher SNR) with the same x-ray techniques, but the spatial resolution in the slice thickness dimension is reduced. Thin slices improve spatial resolution in the thickness dimension and reduce partial volume averaging. For thin slices, the mAs of the study protocol usually is increased to partially compensate for the loss of x-ray photons resulting from the collimation.

It is common to think of a CT image as literally having the geometry of a slab of tissue, but this is not actually the case. The contrast of a small (e.g., 0.5 mm), highly attenuating ball bearing is greater if the bearing is in the center of the CT slice, and the contrast decreases as the bearing moves toward the edges of the slice (Fig. 13-17). This effect describes the *slice sensitivity profile*. For single detector array scanners, the shape of the slice sensitivity profile is a consequence of the finite width of the x-ray focal spot, the penumbra of the collimator, the fact that the image is computed from a number of projection angles encircling the patient, and other minor factors. Furthermore, helical scans have a slightly broader slice sensitivity profile due to translation of the patient during the scan. The *nominal slice thickness* is that which is set on the scanner control panel. Conceptually, the nominal slice is thought of as having a rectangular slice sensitivity profile.

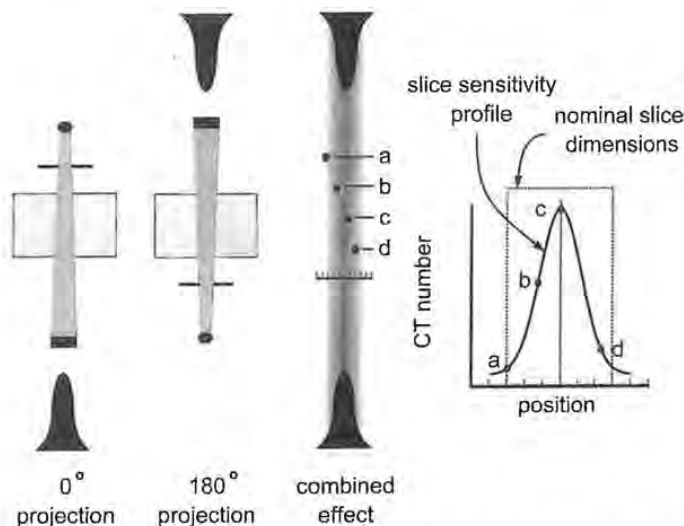


FIGURE 13-17. The slice sensitivity profile is a result of detector width; focal spot size and distribution; collimator penumbra, and the combined influence of all the projection angles. The nominal slice dimensions convey that the computed tomographic (CT) image is of a slab of tissue of finite thickness; however, the same object at different locations in the slice volume (e.g., positions A, B, C, and D) will produce different CT numbers in the image. Consequently, when a small object is placed in the center of the CT slice, it produces greater contrast from background (greater difference in CT number) than when the same object is positioned near the edge of the slice volume.

Slice Thickness: Multiple Detector Array Scanners

The slice thickness of multiple detector array CT scanners is determined not by the collimation, but rather by the width of the detectors in the slice thickness dimension. The width of the detectors is changed by *binning* different numbers of individual detector elements together—that is, the electronic signals generated by adjacent detector elements are electronically summed. Multiple detector arrays can be used both in conventional axial scanning and in helical scanning protocols. In axial scanning (i.e., with no table movement) where, for example, four detector arrays are used, the width of the two center detector arrays almost completely dictates the thickness of the slices. For the two slices at the edges of the scan (detector arrays 1 and 4 of the four active detector arrays), the inner side of the slice is determined by the edge of the detector, but the outer edge is determined either by the collimator penumbra or the outer edge of the detector, depending on collimator adjustment (Fig. 13-18). With a multiple detector array scanner in helical mode, each detector array contributes to every reconstructed image, and therefore the slice sensitivity profile for each detector array needs to be similar to reduce artifacts. To accommodate this condition, it is typical to adjust the collimation so that the focal spot–collimator blade penumbra falls outside the edge detectors. This causes the radiation dose to be a bit higher (especially for small slice widths) in multislice scanners, but it reduces artifacts by equalizing the slice sensitivity profiles between the detector arrays.

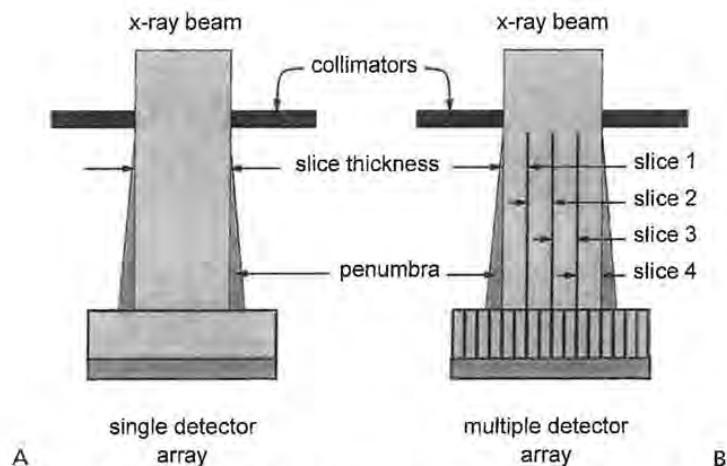


FIGURE 13-18. A: In a single detector array computed tomography (CT) scanner, the width of the CT slice is determined by the collimator. With single detector array scanners, the collimator width is always smaller than the maximum detector width. **B:** Multiple detector array CT scanners acquire several slices simultaneously (four slices shown here), and the x-ray beam collimation determines the outer two edges of the acquired slices. Slice thickness is changed in the single detector array by mechanically changing the collimator position. Slice thickness is determined in the multiple detector array scanner by binning different numbers of detector subunits together and by physically moving the collimator to the outer edges of all four slices.

Detector Pitch and Collimator Pitch

Pitch is a parameter that comes to play when helical scan protocols are used. In a helical CT scanner with one detector array, the pitch is determined by the collimator (collimator pitch) and is defined as:

$$\text{Collimator pitch} = \frac{\text{table movement (mm) per 360-degree rotation of gantry}}{\text{collimator width (mm) at isocenter}}$$

It is customary in CT to measure the collimator and detector widths at the isocenter of the system. The collimator pitch (Fig. 13-19) represents the traditional notion of *pitch*, before the introduction of multiple detector array CT scanners. Pitch is an important component of the scan protocol, and it fundamentally influences radiation dose to the patient, image quality, and scan time. For single detector array scanners, a pitch of 1.0 implies that the number of CT views acquired, when averaged over the long axis of the patient, is comparable to the number acquired with contiguous axial CT. A pitch of less than 1.0 involves overscanning, which may result in some slight improvement in image quality and a higher radiation dose to the patient. CT manufacturers have spent a great deal of developmental effort in optimizing scan protocols for pitches greater than 1.0, and pitches up to 1.5 are commonly used. Pitches with values greater than 1.0 imply some degree of partial scanning along the long axis of the patient. The benefit is faster scan time, less patient motion, and, in some circumstances, use of a smaller volume of contrast agent. Although CT acquisitions around 360 degrees are typical for images of the highest fidelity, the minimum requirement to produce an adequate CT image is a scan of 180 degrees plus the fan angle. With fan angles commonly at about 60 degrees, this means that, at a minimum, $(180 + 60)/360$, or 0.66, of the full circle is required.

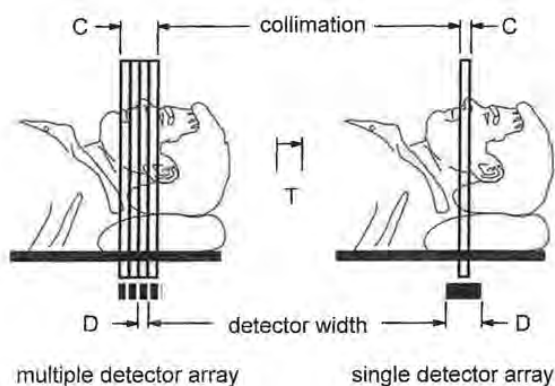


FIGURE 13-19. With a single detector array computed tomographic (CT) scanner, the collimation width is always narrower than the maximum single detector width. A multiple detector array scanner uses collimation that is always wider than a single detector array width. Letting T represent the table translation distance during a 360-degree rotation of the gantry, C would be the collimation width and D would be the detector width. Collimator pitch is defined as T/C , and detector width is defined by T/D . For a multiple detector array CT scanner with four detector arrays, a collimator pitch of 1.0 is equal to a detector pitch of 4.0.

This implies that the upper limit on pitch should be about $1/0.66$, or 1.5, because 66% of the data in a 1.5-pitch scan remains contiguous.

Scanners that have multiple detector arrays require a different definition of pitch. The collimator pitch defined previously is still valid, and collimator pitches range between 0.75 and 1.5, as with single detector array scanners. The *detector pitch* (see Fig. 13-19) is also a useful concept for multiple detector array scanners, and it is defined as:

$$\text{Detector pitch} = \frac{\text{table movement (mm) per 360-degree rotation of gantry}}{\text{detector width (mm)}}$$

The need to define detector pitch and collimator pitch arises because beam utilization between single and multiple detector array scanners is different. For a multiple detector array scanner with N detector arrays, the collimator pitch is as follows:

$$\text{Collimator pitch} = \frac{\text{Detector pitch}}{N}$$

For scanners with four detector arrays, detector pitches running from 3 to 6 are used. A detector pitch of 3 for a four-detector array scanner is equivalent to a collimator pitch of 0.75 ($3/4$), and a detector pitch of 6 corresponds to a collimator pitch of 1.5 ($6/4$).

13.5 TOMOGRAPHIC RECONSTRUCTION

Rays and Views: The Sinogram

The data acquired for one CT slice can be displayed before reconstruction. This type of display is called a *sinogram* (Fig. 13-20). Sinograms are not used for clinical purposes, but the concepts that they embody are interesting and relevant to understanding tomographic principles. The horizontal axis of the sinogram corresponds to the different rays in each projection. For a third-generation scanner, for example, the horizontal axis of the sinogram corresponds to the data acquired at one instant in time along the length of the detector array. A bad detector in a third-generation scanner would show up as a vertical line on the sinogram. The vertical axis in the sinogram represents each projection angle. A state-of-the-art CT scanner may acquire approximately 1,000 views with 800 rays per view, resulting in a sinogram that is 1,000 pixels tall and 800 pixels wide, corresponding to 800,000 data points. The number of data points (rays/view $\times$ views) acquired by the CT scanner has some impact on the final image quality. For example, first- and second-generation scanners used 28,800 and 324,000 data points, respectively. A modern 512×512 circular CT image contains about 205,000 image pixels, and therefore the ratio between raw data points and image pixels is in the range of about 3.2 to 3.9 ($800,000/205,000 = 3.9$).

The number of rays used to reconstruct a CT image has a profound influence on the radial component of spatial resolution, and the number of views affects the circumferential component of the resolution (Fig. 13-21). CT images of a simulated object reconstructed with differing numbers of rays (Fig. 13-22) show that reducing the ray sampling results in low-resolution, blurred images. CT images of the simulated object reconstructed with differing numbers of views

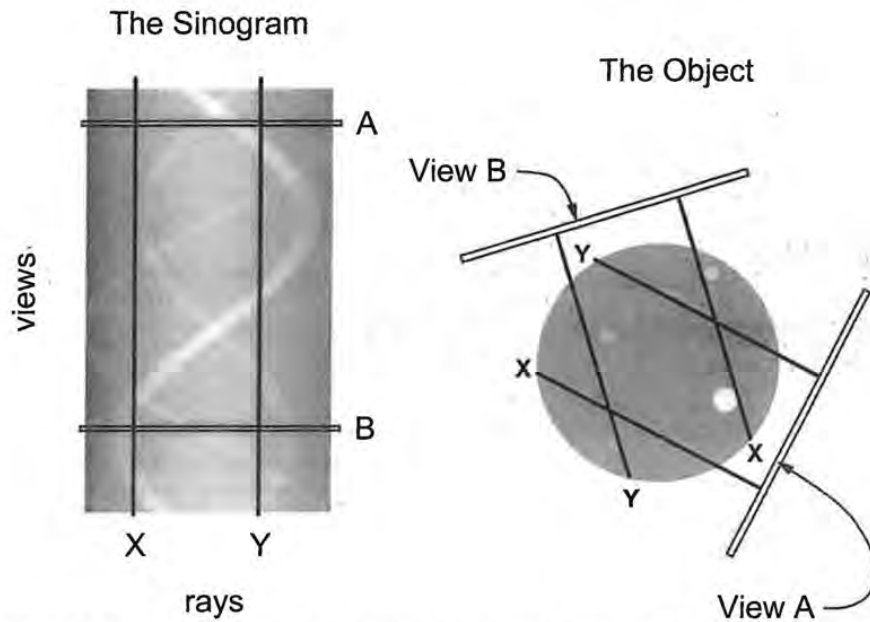


FIGURE 13-20. The sinogram is an image of the raw data acquired by a computed tomographic (CT) scanner before reconstruction. Here, the rays are plotted horizontally and the views are shown on the vertical axis. The sinogram illustrated corresponds to the circular object shown with four circular components. The two views marked A and B on the sinogram correspond to the two views illustrated with the object. During the 360-degree CT acquisition of a particular object, the position of the ray corresponding to that object varies sinusoidally as a function of the view angle. Objects closer to the edge of the field of view produce a sinusoid of high amplitude, and objects closer to the center of rotation have reduced sinusoidal amplitude. An object located exactly at the center of rotation of the gantry produces a shadow that appears vertical on the sinogram.

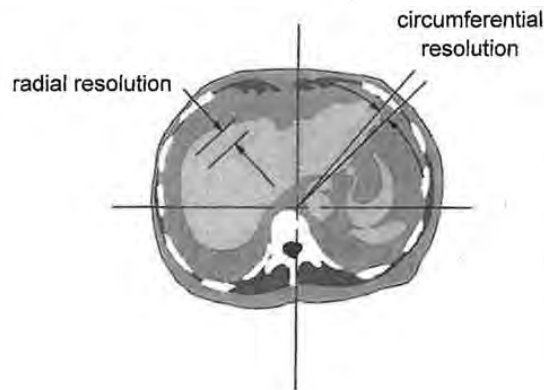


FIGURE 13-21. Unlike screen-film radiography, where the resolution of the image is largely independent of position, the spatial resolution in computed tomography, because of the reconstruction process, has both a radial and circumferential component. The radial component is governed primarily by the spacing and width of the ray data. The circumferential resolution is governed primarily by the number of views used to reconstruct the image.

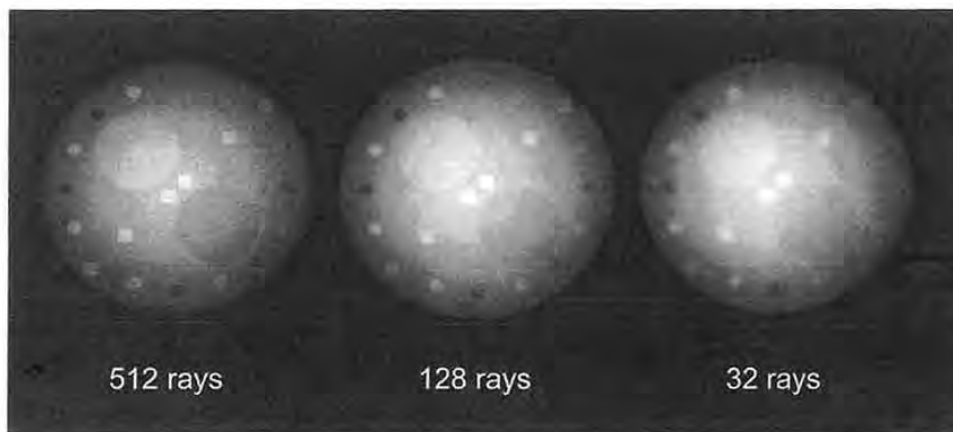


FIGURE 13-22. An image of a test object is shown, reconstructed with 960 views in each image, but with the number of rays differing as indicated. As the number of rays is reduced, the detector aperture was made wider to compensate. With only 32 rays, the detector aperture becomes quite large and the image is substantially blurred as a result.

(Fig. 13-23) show the effect of using too few angular views (*view aliasing*). The sharp edges (high spatial frequencies) of the bar patterns produce radiating artifacts that become more apparent near the periphery of the image. The view aliasing is mild with 240 views and becomes pronounced with 60 views. The presence of high-frequency objects (sharp edges) in the image exacerbates the impact of view aliasing.



FIGURE 13-23. Each of the three images was reconstructed using 512 rays but differing numbers of views. In going from 960 to 240 views, a slight degree of view aliasing is seen (**center image**). When only 60 views are used (**right image**), the view aliasing is significant. The presence of the aliasing is exacerbated by objects with sharp edges, such as the line pair phantom in the image.

Preprocessing the Data

The raw data acquired by a CT scanner is preprocessed before reconstruction. Numerous preprocessing steps are performed. Electronic detector systems produce a digitized data set that is easily processed by the computer. Calibration data determined from air scans (performed by the technologist or service engineer periodically) provide correction data that are used to adjust the electronic gain of each detector in the array. Variation in geometric efficiencies caused by imperfect detector alignments is also corrected. For example, in fourth-generation scanners (see Fig. 13-9), the x-ray source rotates in an arc around each of the detectors in the array. As the angle and the distance between the detector and the x-ray tube change over this arc, the geometric efficiency of x-ray detection changes. (Similarly, the intensity of sunlight shining on one's palm changes as the wrist is rotated.) These geometric efficiencies are measured during calibration scans and stored in the computer; they are corrected in the preprocessing steps.

After the digital data are calibrated, the logarithm of the signal is computed by the following equation:

$$\ln(I_0/I_t) = \mu t$$

where I_0 is the reference data, I_t is the data corresponding to each ray in each view, t is the patient's overall thickness (in centimeters) for that ray, and μ is the linear attenuation coefficient (in cm^{-1}). Note that the data used to calculate the images is linear with respect to the linear attenuation coefficient, μ . The linear attenuation coefficient is determined by the composition and density of the tissue inside each voxel in the patient. Whereas μ is the total attenuation coefficient for each ray (Fig. 13-24), it can be broken down into its components in each small path length Δt :

$$\mu t = \mu_1 \Delta t + \mu_2 \Delta t + \mu_3 \Delta t + \dots + \mu_n \Delta t = \Delta t [\mu_1 + \mu_2 + \mu_3 + \dots + \mu_n]$$

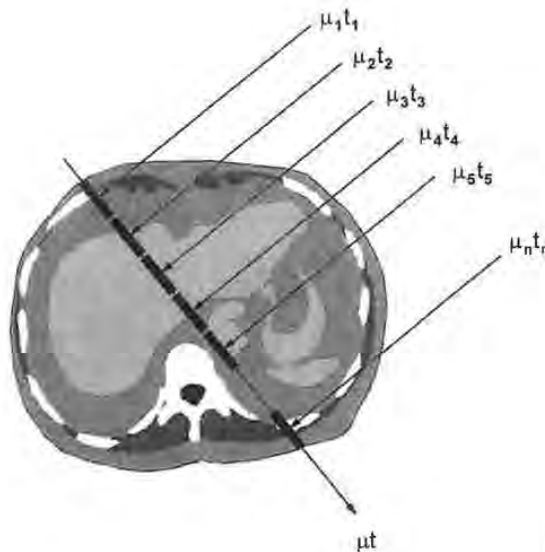


FIGURE 13-24. The measured value of a single ray is equivalent to the attenuation path (t) times the linear attenuation coefficient (μ). The single measurement of μt can be broken up into a series of measurements $\mu_1 t_1$, $\mu_2 t_2$, ..., $\mu_n t_n$, as shown.

The constant Δt factors out, resulting in the following:

$$\mu = \mu_1 + \mu_2 + \mu_3 + \dots + \mu_n$$

Therefore, the reconstructed value in each pixel is the linear attenuation coefficient for the corresponding voxel.

Interpolation (Helical)

Helical CT scanning produces a data set in which the x-ray source has traveled in a helical trajectory around the patient. Present-day CT reconstruction algorithms assume that the x-ray source has negotiated a circular, not a helical, path around the patient. To compensate for these differences in the acquisition geometry, before the actual CT reconstruction the helical data set is interpolated into a series of planar image data sets, as illustrated in Fig. 13-25. Although this interpolation represents an additional step in the computation, it also enables an important feature. With conventional axial scanning, the standard is to acquire contiguous images, which abut one another along the cranial-caudal axis of the patient. With helical scanning, however, CT images can be reconstructed at any position along the length of the scan to within $(\frac{1}{2})$ (pitch) (slice thickness) of each edge of the scanned volume. Helical scanning allows the production of additional overlapping images with *no additional dose* to the patient. Figure 13-26 illustrates the added clinical importance of using interleaved images. The sensitivity of the CT image to objects not centered in the voxel is reduced (as quantified by the slice sensitivity profile), and therefore subtle lesions, which lay between two contiguous images, may be missed. With helical CT scanning, interleaved reconstruction allows the placement of additional images along the patient, so that the clinical examination is almost uniformly sensitive to subtle abnormalities. Interleaved reconstruction adds no additional radiation dose to the patient, but additional time is required to reconstruct the images. Although an increase in the

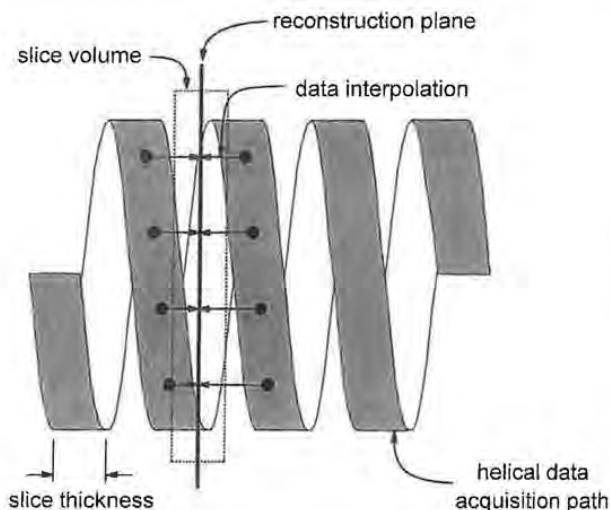


FIGURE 13-25. During helical acquisition, the data are acquired in a helical path around the patient. Before reconstruction, the helical data are interpolated to the reconstruction plane of interest. Interpolation is essentially a weighted average of the data from either side of the reconstruction plane, with slightly different weighting factors used for each projection angle.

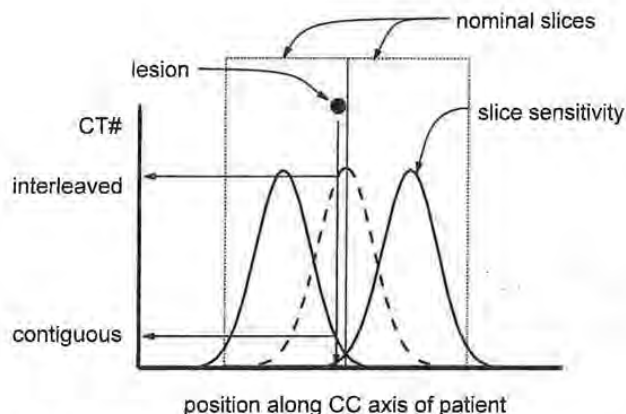


FIGURE 13-26. This figure illustrates the value of interleaved reconstruction. The nominal slice for contiguous computed tomographic (CT) images is illustrated conceptually as two adjacent rectangles; however, the sensitivity of each CT image is actually given by the slice sensitivity profile (*solid lines*). A lesion that is positioned approximately between the two CT images (*black circle*) produces low contrast (i.e., a small difference in CT number between the lesion and the background) because it corresponds to low slice sensitivity. With the use of interleaved reconstruction (*dashed line*), the lesion intersects the slice sensitivity profile at a higher position, producing higher contrast.

image count would increase the interpretation time for traditional side-by-side image presentation, this concern will ameliorate as more CT studies are read by radiologists at computer workstations.

It is important not to confuse the ability to reconstruct CT images at short intervals along the helical data set with the axial resolution itself. The slice thickness (governed by collimation with single detector array scanners and by the detector width in multislice scanners) dictates the actual spatial resolution along the long axis of the patient. For example, images with 5-mm slice thickness can be reconstructed every 1 mm, but this does not mean that 1-mm spatial resolution is achieved. It simply means that the images are sampled at 1-mm intervals. To put the example in technical terms, the *sampling pitch* is 1 mm but the *sampling aperture* is 5 mm. In practice, the use of interleaved reconstruction much beyond a 2:1 interleave yields diminishing returns, except for multiplanar reconstruction (MPR) or 3D rendering applications.

Simple Backprojection Reconstruction

Once the image raw data have been preprocessed, the final step is to use the planar projection data sets (i.e., the preprocessed sinogram) to reconstruct the individual tomographic images. *Iterative reconstruction techniques* are discussed in Chapter 22 and will not be discussed here because they are not used in clinical x-ray CT. As a basic introduction to the reconstruction process, consider Fig. 13-27. Assume that a very simple 2×2 "image" is known only by the projection values. Using algebra ("N equations in M unknowns"), one can solve for the image values in the simple case of a 4-pixel image. A modern CT image contains approx-

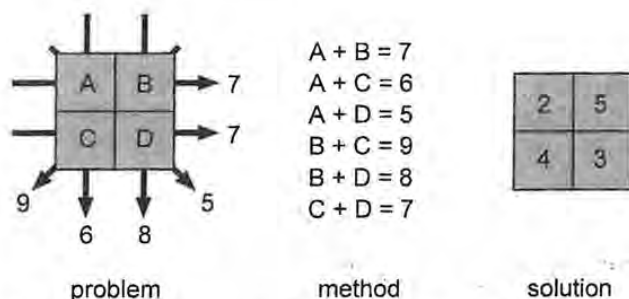


FIGURE 13-27. The mathematical problem posed by computed tomographic (CT) reconstruction is to calculate image data (the pixel values—A, B, C, and D) from the projection values (arrows). For the simple image of four pixels shown here, algebra can be used to solve for the pixel values. With the six equations shown, using substitution of equations, the solution can be determined as illustrated. For the larger images of clinical CT, algebraic solutions become unfeasible, and filtered backprojection methods are used.

imately 205,000 pixels (the circle within a 512×512 matrix) or “unknowns,” and each of the 800,000 projections represent an individual equation. Solving this kind of a problem is beyond simple algebra, and backprojection is the method of choice.

Simple backprojection is a mathematical process, based on trigonometry, which is designed to emulate the acquisition process in reverse. Each ray in each view rep-

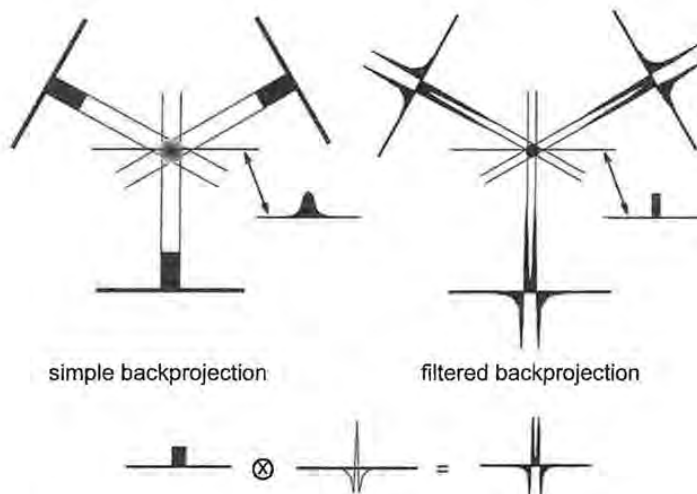


FIGURE 13-28. Simple backprojection is shown on the left; only three views are illustrated, but many views are actually used in computed tomography. A profile through the circular object, derived from simple backprojection, shows a characteristic $1/r$ blurring. With filtered backprojection, the raw projection data are convolved with a convolution kernel and the resulting projection data are used in the backprojection process. When this approach is used, the profile through the circular object demonstrates the crisp edges of the cylinder, which accurately reflects the object being scanned.

resents an individual measurement of μ . In addition to the value of μ for each ray, the reconstruction algorithm also “knows” the acquisition angle and position in the detector array corresponding to each ray. Simple backprojection starts with an empty image matrix (an image with all pixels set to zero), and the μ value from each ray in all views is smeared or backprojected onto the image matrix. In other words, the value of μ is added to each pixel in a line through the image corresponding to the ray’s path.

Simple backprojection comes very close to reconstructing the CT image as desired. However, a characteristic $1/r$ blurring (see later discussion) is a byproduct of simple backprojection. Imagine that a thin wire is imaged by a CT scanner perpendicular to the image plane; this should ideally result in a small point on the image. Rays not running through the wire will contribute little to the image ($\mu = 0$). The backprojected rays, which do run through the wire, will converge at the position of the wire in the image plane, but these projections run from one edge of the reconstruction circle to the other. These projections (i.e., lines) will therefore “radiate” geometrically in all directions away from a point input. If the image gray scale is measured as a function of distance away from the center of the wire, it will gradually diminish with a $1/r$ dependency, where r is the distance away from the point. This phenomenon results in a blurred image of the actual object (Fig. 13-28) when simple backprojection is used. A filtering step is therefore added to correct this blurring, in a process known as *filtered backprojection*.

Filtered Backprojection Reconstruction

In filtered backprojection, the raw view data are mathematically filtered before being backprojected onto the image matrix. The filtering step mathematically reverses the image blurring, restoring the image to an accurate representation of the object that was scanned. The mathematical filtering step involves *convolving* the projection data with a convolution *kernel*. Many convolution kernels exist, and different kernels are used for varying clinical applications such as soft tissue imaging or bone imaging. A typical convolution kernel is shown in Fig. 13-28. The kernel refers to the shape of the filter function in the spatial domain, whereas it is common to perform (and to think of) the filtering step in the frequency domain. Much of the nomenclature concerning filtered backprojection involves an understanding of the frequency domain (which was discussed in Chapter 10). The Fourier transform (FT) is used to convert a function expressed in the spatial domain (millimeters) into the spatial frequency domain (cycles per millimeter, sometimes expressed as mm^{-1}); the inverse Fourier transform (FT^{-1}) is used to convert back. Convolution is an integral calculus operation and is represented by the symbol $\otimes$. Let $p(x)$ represent projection data (in the spatial domain) at a given angle ($p(x)$ is just one horizontal line from a sinogram; see Fig. 13-20), and let $k(x)$ represent the spatial domain kernel as shown in Fig. 13-28. The filtered data in the spatial domain, $p'(x)$, is computed as follows:

$$p'(x) = p(x) \otimes k(x)$$

The difference between filtered backprojection and simple backprojection is the mathematical filtering operation (convolution) shown in this equation above. In filtered backprojection, $p'(x)$ is backprojected onto the image matrix, whereas in

simple backprojection, $p(x)$ is backprojected. The equation can also be performed, quite exactly, in the frequency domain:

$$p'(x) = \text{FT}^{-1}\{\text{FT}[p(x)] \times K(f)\}$$

where $K(f) = \text{FT}[k(x)]$, the kernel in the frequency domain. This equation states that the convolution operation can be performed by Fourier transforming the projection data, *multiplying* (not convolving) this by the frequency domain kernel ($K(f)$), and then applying the inverse Fourier transform on that product to get the filtered data, ready to be backprojected.

Various convolution filters can be used to emphasize different characteristics in the CT image. Several filters, shown in the frequency domain, are illustrated in Fig. 13-29, along with the reconstructed CT images they produced. The Lak filter, named for Dr. Lakshminarayanan, increases the amplitude linearly as a function of frequency and is also called a *ramp filter*. The $1/r$ blurring function in the spatial domain becomes a $1/f$ blurring function in the frequency domain; therefore, multiplying by the Lak filter, where $L(f) = f$, exactly compensates the unwanted $1/f$ blurring, because $1/f \times f = 1$, at all f . This filter works well when there is no noise

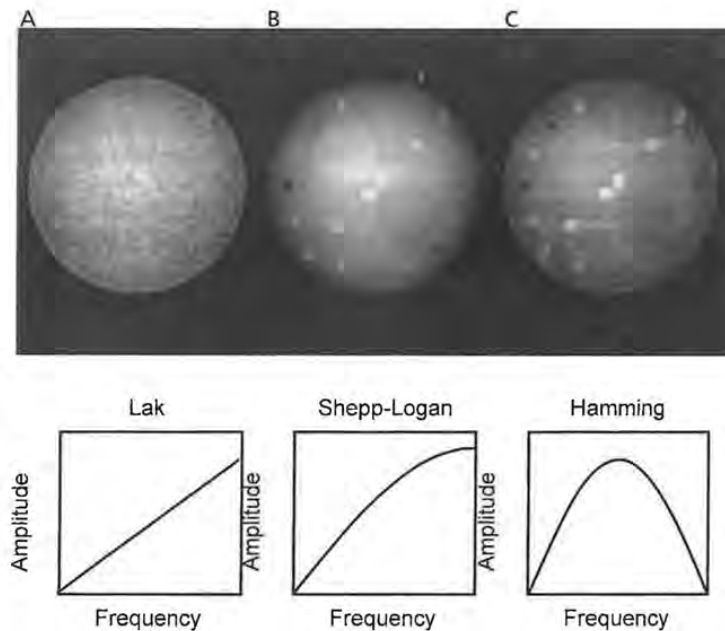


FIGURE 13-29. Three computed tomography (CT) reconstruction filters are illustrated. The plot of the reconstruction filter is shown below, with the image it produced above. **(A)** The Lak is the ideal reconstruction filter in the absence of noise. However, with the quantum noise characteristic in x-ray imaging including CT, the high-frequency components of the noise are amplified and the reconstructed image is consequently very noisy. **(B)** The Shepp-Logan filter is similar to the Lak filter, but roll-off occurs at the higher frequencies. This filter produces an image that is a good compromise between noise and resolution. **(C)** The Hamming filter has extreme high frequency roll-off.

in the data, but in x-ray images there is always x-ray quantum noise, which tends to be more noticeable in the higher frequencies. The Lak filter, as shown in Fig. 13-29A, produces a very noisy CT image. The Shepp-Logan filter in Fig. 13-29B is similar to the Lak filter but incorporates some *roll-off* at higher frequencies, and this reduction in amplification at the higher frequencies has a profound influence in terms of reducing high-frequency noise in the final CT image. The Hamming filter (see Fig. 13-29C) has an even more pronounced high-frequency roll-off, with better high-frequency noise suppression.

Bone Kernels and Soft Tissue Kernels

The reconstruction filters derived by Lakshminarayanan, Shepp and Logan, and Hamming provide the mathematical basis for CT reconstruction filters. In clinical CT scanners, the filters have more straightforward names, and terms such as “bone filter” and “soft tissue filter” are common among CT manufacturers. The term *kernel* is also used. Bone kernels have less high-frequency roll-off and hence accentuate higher frequencies in the image at the expense of increased noise. CT images of bones typically have very high contrast (high signal), so the SNR is inherently quite good. Therefore, these images can afford a slight decrease in SNR ratio in return for sharper detail in the bone regions of the image.

For clinical applications in which high spatial resolution is less important than high contrast resolution—for example, in scanning for metastatic disease in the liver—soft tissue reconstruction filters are used. These kernels have more roll-off at higher frequencies and therefore produce images with reduced noise but lower spatial resolution. The resolution of the images is characterized by the modulation transfer function (MTF); this is discussed in detail in Chapter 10. Figure 13-30 demonstrates two MTFs for state-of-the-art CT scanners. The high-resolution MTF corresponds to use of the bone filter at small FOV, and the standard resolution corresponds to images produced with the soft tissue filter at larger FOV.

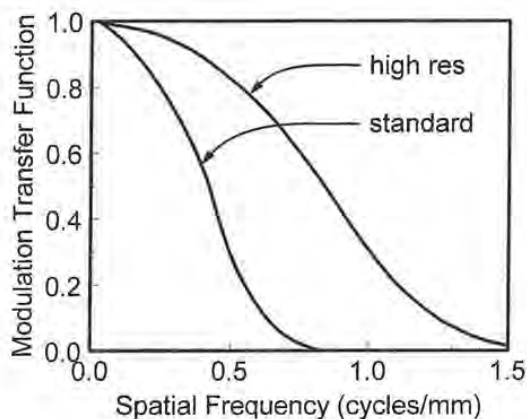


FIGURE 13-30. The modulation transfer function (MTF) for standard and high-resolution clinical computed tomographic scanners are shown. The standard MTF is for a large field of view and a soft tissue reconstruction filter. The high-resolution MTF corresponds to a small field of view and a bone reconstruction filter.

The units for the x-axis in Fig. 13-30 are in cycles per millimeter, and the cut-off frequency is approximately 1.0 cycles/mm. This cutoff frequency is similar to that of fluoroscopy, and it is five to seven times lower than in general projection radiography. CT manufacturers have adapted the unit *cycle/cm*—for example, 1.2 cycles/mm = 12 cycles/cm.

CT Numbers or Hounsfield Units

After CT reconstruction, each pixel in the image is represented by a high-precision floating point number that is useful for computation but less useful for display. Most computer display hardware makes use of integer images. Consequently, after CT reconstruction, but before storing and displaying, CT images are normalized and truncated to integer values. The number $CT(x,y)$ in each pixel, (x,y) , of the image is converted using the following expression:

$$CT(x,y) = 1,000 \frac{\mu(x,y) - \mu_{\text{water}}}{\mu_{\text{water}}}$$

where $\mu(x,y)$ is the floating point number of the (x,y) pixel before conversion, μ_{water} is the attenuation coefficient of water, and $CT(x,y)$ is the CT number (or Hounsfield unit) that ends up in the final clinical CT image. The value of μ_{water} is about 0.195 for the x-ray beam energies typically used in CT scanning. This normalization results in CT numbers ranging from about -1,000 to +3,000, where -1,000 corresponds to air, soft tissues range from -300 to -100, water is 0, and dense bone and areas filled with contrast agent range up to +3,000.

What do CT numbers correspond to *physically* in the patient? CT images are produced with a highly filtered, high-kV x-ray beam, with an average energy of about 75 keV. At this energy in muscle tissue, about 91% of x-ray interactions are Compton scatter. For fat and bone, Compton scattering interactions are 94% and 74%, respectively. Therefore, CT numbers and hence CT images derive their contrast mainly from the physical properties of tissue that influence Compton scatter. Density (g/cm^3) is a very important discriminating property of tissue (especially in lung tissue, bone, and fat), and the linear attenuation coefficient, μ , tracks linearly with density. Other than physical density, the Compton scatter cross section depends on the electron density (ρ_e) in tissue: $\rho_e = NZ/A$, where N is Avogadro's number (6.023×10^{23} , a constant), Z is the atomic number, and A is the atomic mass of the tissue. The main constituents of soft tissue are hydrogen ($Z = 1$, $A = 1$), carbon ($Z = 6$, $A = 12$), nitrogen ($Z = 7$, $A = 14$), and oxygen ($Z = 8$, $A = 16$). Carbon, nitrogen, and oxygen all have the same Z/A ratio of 0.5, so their electron densities are the same. Because the Z/A ratio for hydrogen is 1.0, the relative abundance of hydrogen in a tissue has some influence on CT number. Hydrogenous tissue such as fat is well visualized on CT. Nevertheless, density (g/cm^3) plays the dominant role in forming contrast in medical CT.

CT numbers are *quantitative*, and this property leads to more accurate diagnosis in some clinical settings. For example, pulmonary nodules that are calcified are typically benign, and the amount of calcification can be determined from the CT image based on the mean CT number of the nodule. Measuring the CT number of a single pulmonary nodule is therefore common practice, and it is an important part of the diagnostic work-up. CT scanners measure bone density with good accuracy, and when phantoms are placed in the scan field along with the patient, quan-

titative CT techniques can be used to estimate bone density, which is useful in assessing fracture risk. CT is also quantitative in terms of linear dimensions, and therefore it can be used to accurately assess tumor volume or lesion diameter.

Computed Tomographic Fluoroscopy Reconstruction

In CT fluoroscopy, the scanner provides pseudo-real time tomographic images that are most commonly used for guidance during biopsies. CT fluoroscopy is similar to conventional CT, with a few added technologic nuances. Like conventional fluoroscopy, CT fluoroscopy provides an image sequence over the same region of tissue, and therefore the radiation dose from repetitive exposure to that tissue is a concern. To address this problem, CT scanners are capable of a CT fluoroscopic mode that uses a low milliamperage setting, typically 20 to 50 mA, whereas 150 to 400 mA is typical for conventional CT imaging.

To provide real-time CT images, special computer hardware is required. Each fluoroscopic CT frame in an imaging sequence is not unique from the preceding CT frame. For example, to achieve 6 CT images per second with a 1-second gantry rotation, the angle scanned during each frame time of $1/6$ second is 60 degrees ($360 \text{ degrees} / 6 = 60 \text{ degrees}$). For CT reconstruction using a full 360-degree projection set, each CT fluoroscopic image uses 17% ($1/6$) new information and 83% ($5/6$) old information (Fig. 13-31). The image acquisition time is still 1 second in this example, but the image is updated 6 times per second. This clever approach provides increased temporal resolution given the physical constraints of backprojection algorithms and the hardware constraints of current technology.

The full data set for each CT fluoroscopic image is not completely re-reconstructed during each frame interval. Using the example of 6 frames per second, after 1 second of operation, consider what occurs in the next $1/6$ second (167 msec): The scanner completes a 60-degree arc, and the projection data acquired during that 167

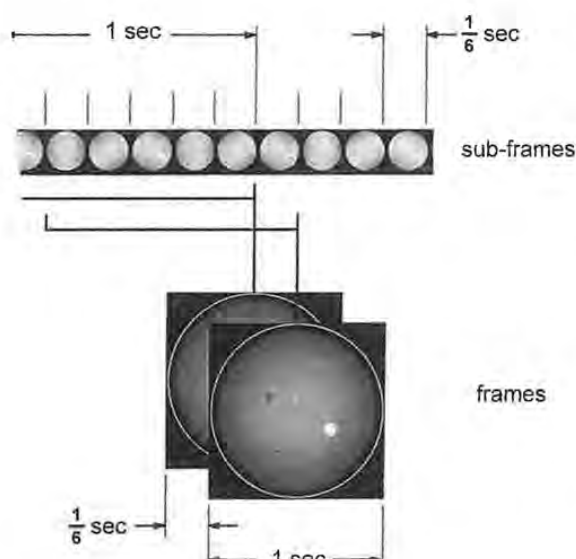


FIGURE 13-31. Data handling in computed tomographic fluoroscopy is illustrated, here using the example of acquisition at 6 frames per second and a 1-second gantry rotation time. During each $1/6$ second, the CT scanner acquires 60 degrees of projection data. Each of these subframes can be filtered and backprojected as shown. A full-frame image is produced by adding six subframes together. Six CT frames are produced each second; however, each CT image corresponds to a scan time of 1 second. For each subsequent CT frame, the newest subframe and the five previous subframes are

msec is preprocessed, mathematically filtered, and backprojected to create a CT subframe. This new CT subframe (an image) is added (pixel by pixel) to the five previous subframes, creating a complete CT frame ready for display. Therefore, every 167 msec, a new subframe is created and an old subframe is discarded, so that the most recent six subframes are summed to produce the displayed CT frame. Although 6 frames per second was used in this example, higher frame rates are clinically available.

13.6 DIGITAL IMAGE DISPLAY

Once the CT images from a patient's examination have been reconstructed, the image data must be conveyed to the physician for review and diagnosis. There are some basic postprocessing techniques that are applied to all CT images (i.e., windowing and leveling). For CT studies that have contiguous or interleaving images over a length of the patient, the volume data set lends itself to other forms of postprocessing.

Windowing and Leveling

CT images typically possess 12 bits of gray scale, for a total of 4,096 shades of gray (CT numbers range from $-1,000$ to $+3,095$). Electronic display devices such as computer monitors have the ability to display about 8 bits (256 shades of gray), and laser imagers used for "filming" CT studies have about 8 bits of display fidelity as well. The human eye has incredible dynamic range for adapting to *absolute* light intensities, from moonless nights to desert sunshine. Much of this dynamic range is a result of the eye's ability to modulate the diameter of the light-collecting aperture (the pupil). However, for resolving *relative* differences in gray scale at fixed luminance levels (fixed pupil diameter), as in viewing medical images, the human eye has limited ability (30 to 90 shades of gray), and 6 to 8 bits is considered sufficient for image display.

The 12-bit CT images must be reduced to 8 bits to accommodate most image display hardware. The most common way to perform this postprocessing task (which nondestructively adjusts the image contrast and brightness), is to *window and level* the CT image (Fig. 13-32). The window width (W) determines the contrast of the image, with narrower windows resulting in greater contrast. The level (L) is the CT number at the center of the window. Selection of the values of L and W determines the inflection points P_1 and P_2 , as shown in Fig. 13-32, where $P_1 = L - \frac{1}{2}W$ and $P_2 = L + \frac{1}{2}W$. All the CT numbers lower than P_1 will be saturated to uniform black, and all the CT numbers above P_2 will be saturated to uniform white, and in these saturated areas all image information is lost. It is routine to display CT images using several (two or three) different window and level settings for each image.

Multiplanar Reconstruction

A stack of axial CT images represents anatomic information in three dimensions; however, the axial display is the most common. For studying anatomic features that run along the cranial-caudal dimension of the body, such as the aorta or spinal cord,

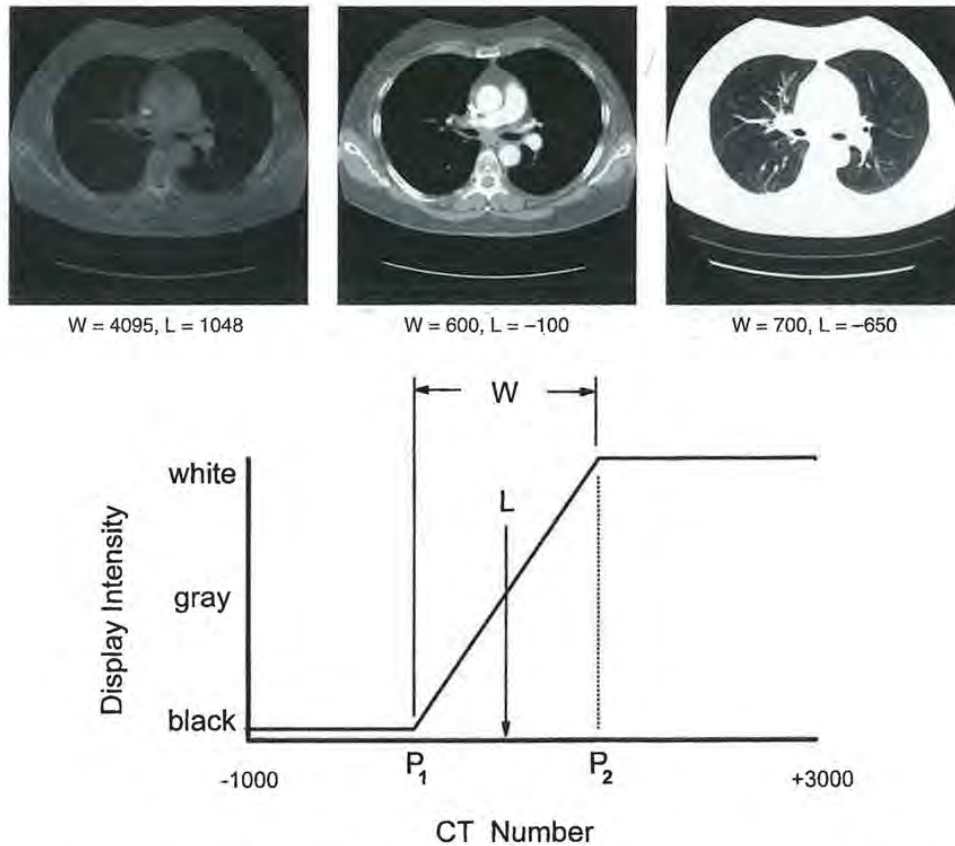


FIGURE 13-32. The concept of how the window and level values are used to manipulate the contrast of the CT image is illustrated. The *level* is the CT number corresponding to the center of the *window*. A narrow window produces a very high contrast image, corresponding to a large slope on the figure. CT numbers below the window (lower than P_1) will be displayed on the image as black; CT numbers above the window (higher than P_2) will be displayed as white. Only CT numbers between P_1 and P_2 will be displayed in a meaningful manner. The thoracic CT images (**top**) illustrate the dramatic effect of changing the window and level settings.

it is sometimes convenient to recompute CT images for display in the sagittal or coronal projections. This is a straightforward technique (Fig. 13-33), and is simply a reformatting of the image data that already exists. Some MPR software packages allow the operator to select a curved reconstruction plane for sagittal or coronal display, to follow the anatomic structure of interest. The spatial resolution of the sagittal or coronal CT images is typically reduced, compared with the axial views. The in-plane pixel dimensions approximate the x-y axis resolution, but the slice thickness limits the z-axis resolution. Sagittal and coronal MPR images combine the x or y dimensions of the CT image with image data along the z-axis, and therefore a mismatch in spatial sampling and resolution occurs. The MPR software compensates for the sampling mismatch by interpolation, so that the MPR images maintain

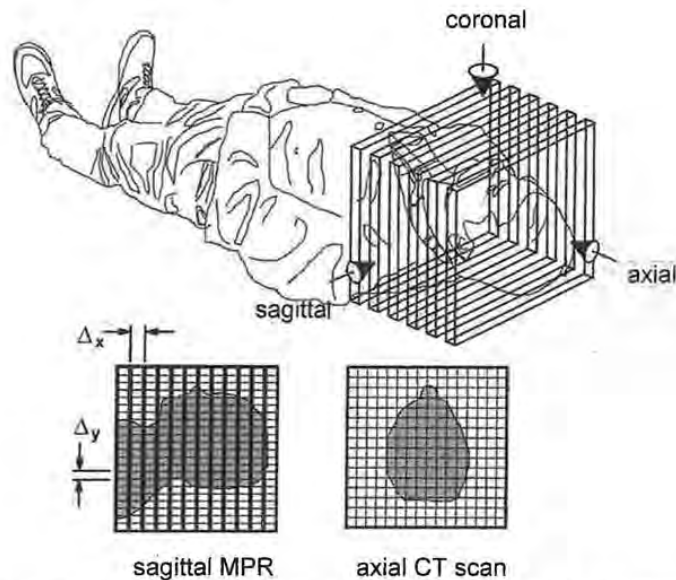


FIGURE 13-33. The coronal, sagittal, and axial projections are illustrated for reference. Both computed tomographic (CT) scans are produced as axial images, and within the image the pixel-to-pixel spacing is small (perhaps 0.5 mm, as a very general number). Coronal or sagittal multiplanar reconstructions (MPRs) can be produced from a series of axial CT scans. For instance, in a sagittal MPR image (**bottom left**), the vertical pixel spacing (Δy) is equivalent to the spacing of the pixels on the axial image. The horizontal spacing (Δx) corresponds to the distance between the reconstructed images. Because Δy is usually much smaller than Δx , it is common to interpolate image data in the horizontal dimension to produce square pixels. For MPR when helical acquisition was used, it is advantageous to reconstruct the images with a significant degree of interleave.

proper aspect ratio (pixel height equals pixel width). However, the system cannot improve the resolution in the z dimension, so MPR images appear a bit blurry along this axis. If it is known before the CT study that MPR images are to be computed, the use of a protocol with thinner CT slices (down to 1 mm) to improve z -axis resolution is warranted.

Three-Dimensional Image Display

Most radiologists are sufficiently facile at viewing series of 2D images to garner a 3D understanding of the patient's anatomy and pathology. In some cases, however, referring physicians prefer to see the CT image data presented in a pseudo-3D manner. Examples would include planning for craniofacial surgery, aortic stent grafting, or radiation treatment. There are two general classes of volume reconstruction techniques: volume rendering techniques and reprojection techniques (Fig. 13-34).

Volume rendering of CT images requires *segmentation*, the identification (by computer alone or with the aid of human intervention) of specific target structures on the 2D image series, before 3D rendering. Segmentation is easiest to perform



FIGURE 13-34. A volume-rendered image and maximum intensity projection (MIP) image of the aorta and renal arteries are illustrated. Volume-rendered images use surface shading techniques to produce the appearance of a three-dimensional image. The MIP display is similar to a projection radiograph of the volume data set. (Images courtesy of Rick Labonte.)

automatically when there is a large difference in CT number between the target structures and adjacent anatomy (e.g., between bone and soft tissue). The simplest segmentation algorithms allow the user to select a CT number, and the algorithm then assumes that the target structure occupies all pixels with CT numbers greater than the threshold value. Segmentation of the CT data set essentially converts the images to one-bit representations: a "0" if the target structure is not in the pixel and a "1" if it is. Once the target structures (e.g., the facial bones) have been accurately segmented, a computer program then calculates ("renders") the theoretical appearance of the structure from a specified angle. The software calculates a number of surfaces from the segmented data set, and this process is sometimes called *surface rendering*. Surfaces closest to the viewer obscure surfaces at greater depth. Depending on the sophistication of the software, single or multiple light sources may be factored into the computation to give the appearance of surface shading, color may be added to identify different target structures (i.e., green tumor, white bones, pink spinal cord), and so on. Volume rendering has reached spectacular technologic sophistication in the movie industry, with *Jurassic Park* being the classic example. Medical imaging software has and will continue to benefit from nonmedical rendering advancements, but it will never meet the same degree of success because of the SNR limitations in the data set. The dinosaurs in *Jurassic Park* were calculated from sophisticated mathematical models with perfect SNR, whereas rendering of cranial-facial bones requires a clinical CT data set with imperfect SNR.

Because of SNR limitations in CT imaging, in some studies segmentation, which is required for volume rendering, can be a time-consuming task requiring substantial human intervention, reducing the appeal of this approach. To address this problem, reprojection techniques that do not require segmentation have been developed for the display of volume data. From a specified viewing angle, images are computed from the CT volume data set which are similar geometrically to radiographic projection images. Reprojection techniques simply use ray-tracing soft-

ware through the volume data set, from the specified viewing angle. The software can display the normalized sum of the CT numbers from all the voxels through which each ray passed. More commonly, the maximum CT number that each ray encounters is displayed, and this mode is called *maximum intensity projection* (MIP). MIP displays are perhaps not as "three-dimensional" as volume-rendered images, but they can be generated quite reproducibly and completely automatically. To augment the 3D appearance of MIP images, often a series of such images are produced at different viewing angles, and a short movie sequence is generated. For example, a sequence rotating the cervical spine in real time could be displayed. The spatial-temporal cues from the real-time display of these movies can be quite revealing, especially for complicated structures.

Stack Mode Viewing

Many radiology departments are moving to a filmless environment, in which the actual diagnostic interpretation is performed by a radiologist who sits at a computer workstation and views *soft copy* images. One approach is to try to emulate the film alternator to the extent possible, with perhaps four or even six high-resolution monitors clustered together to display the entire CT data set simultaneously. Although this approach seeks to emulate an environment with which most radiologists are familiar, four- or six-monitor workstations are expensive and bulky, and this approach is not designed to take advantage of a key attribute of soft copy display—the computer.

An alternative approach is to simplify the computer hardware, to perhaps two high-resolution monitors, and allow the computer to do the work of the radiologist's neck. In stack mode, a single CT image is displayed and the radiologist selects the image in the study, which is displayed by moving the computer mouse. Often the scout view is presented next to the CT image, with the current image location highlighted. It is possible to simultaneously display two CT images at the same cut plane, such as precontrast and postcontrast images, images at different window and level settings, or images from the previous and the current CT examinations. Certainly many permutations of stack mode display exist, but the essential advantage of this approach is that it is interactive: The radiologist interacts with the computer in real time to visualize the image data as he or she interprets the case, following diagnostic clues from slice to slice. As one clinical example of the benefits of stack mode viewing, a common task in CT is to follow arteries from image to image. On an alternator-like, noninteractive display, the viewer must relocate the vessel spatially, shifting the gaze from image to image. In stack mode display, the viewer's eyes are fixed on the vessel while the various images are displayed under mouse control, allowing the viewer's gaze to follow the vessel as it changes position slightly from image to image.

13.7 RADIATION DOSE

Different x-ray modalities address radiation dose in different ways. For example, in chest radiography it is the entrance *exposure* (not the *dose*) that is the commonly quoted comparison figure. In mammography, the *average glandular dose* is the standard measure of dose. The distribution of radiation dose in CT is markedly differ-

ent than in radiography, because of the unique way in which radiation dose is deposited. There are three aspects of radiation dose in CT that are unique in comparison to x-ray projection imaging. First, because a single CT image is acquired in a highly collimated manner, the volume of tissue that is irradiated by the primary x-ray beam is substantially smaller compared with, for example, the average chest radiograph. Second, the volume of tissue that is irradiated in CT is exposed to the x-ray beam from almost all angles during the rotational acquisition, and this more evenly distributes the radiation dose to the tissues in the beam. In radiography, the tissue irradiated by the entrance beam experiences exponentially more dose than the tissue near the exit surface of the patient. Finally, CT acquisition requires a high SNR to achieve high contrast resolution, and therefore the radiation dose to the slice volume is higher because the techniques used (kV and mAs) are higher. As a rough comparison, a typical PA chest radiograph may be acquired with the use of 120 kV and 5 mAs, whereas a thoracic CT image is typically acquired at 120 kV and 200 mAs.

Dose Measurement

Compton scattering is the principal interaction mechanism in CT, so the radiation dose attributable to scattered radiation is considerable, and it can be higher than the radiation dose from the primary beam. Scattered radiation is not confined to the collimated beam profile as primary x-rays are, and therefore the acquisition of a CT slice delivers a considerable dose from scatter to adjacent tissues, outside the primary beam. Furthermore, most CT protocols call for the acquisition of a series of near-contiguous CT slices over the tissue volume under examination. Take, for example, a protocol in which ten 10-mm CT slices are acquired in the abdomen. The tissue in slice 5 will receive both primary and scattered radiation from its acquisition, but it will also receive the scattered radiation dose from slices 4 and 6, and to a lesser extent from slices 3 and 7, and so on (Fig. 13-35).

The *multiple scan average dose* (MSAD) is the standard for determining radiation dose in CT. The MSAD is the dose to tissue that includes the dose attributable to scattered radiation emanating from all adjacent slices. The MSAD is defined as the average dose, at a particular depth from the surface, resulting from a large series

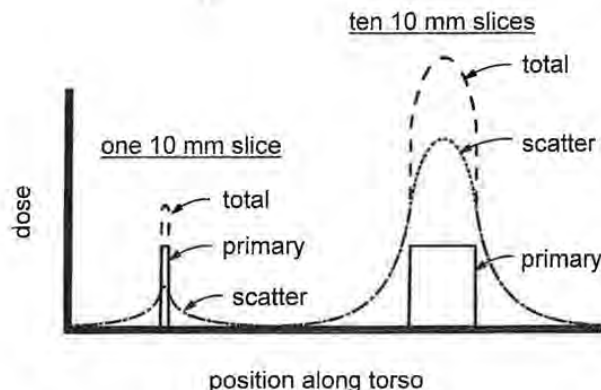


FIGURE 13-35. The radiation dose to the patient is shown as a function of position in the torso along the cranial-caudal dimension. The dose profile for a one 10-mm slice is shown on the left, and the dose profile for ten contiguous 10-mm slices is shown on the right. The contribution (height) of the dose profile from the primary component is the same. However, with a number of adjacent slices, the scattered radiation tails build up and increase the total dose in each CT slice.

of CT slices. The MSAD could be measured directly by placing a small exposure meter at a point in a dose phantom, taking a large series of CT scans of the phantom with the meter in the middle of the slices, and adding the doses from all slices.

An estimate of the MSAD can be accomplished with a single scan by measuring the *CT dose index* (CTDI). It can be shown that the CTDI provides a good approximation to the MSAD when the slices are contiguous. The CTDI measurement protocol seeks to measure the scattered radiation dose from adjacent CT slices in a practical manner. The CTDI is defined by the U.S. Food and Drug Agency (21 CFR § 1020.33) as the radiation dose to any point in the patient including the scattered radiation contribution from 7 CT slices in both directions, for a total of 14 slices. In this text, this definition is referred to as the CTDI_{FDA}. One method to measure the CTDI_{FDA} is with many small thermoluminescent dosimeters (TLDs) placed in holes along 14-slice-thickness increments in a Lucite dose phantom. (TLDs are discussed in Chapters 20 and 23.) A single CT image is acquired at the center of the rod containing the TLDs, and the CTDI_{FDA} is determined from this single CT scan by summing the TLD dose measurements.

Medical physicists usually measure the CTDI with the use of a long (100 mm), thin *pencil ionization chamber*. The pencil chamber is long enough to span the width of 14 contiguous 7-mm CT scans and provides a good estimation of the CTDI_{FDA} for 7- and 8-mm slices. A single CT image is acquired at the center of the pencil chamber, and the CTDI is determined from this single CT scan. To calculate the CTDI, all of the energy deposition along the length of the ion chamber is assigned to the thickness of the CT slice:

$$\text{CTDI} = fX/T \times L$$

where X is the measured air kerma (mGy) or exposure (R) to the pencil ion chamber, f is an air-kerma-to-dose (mGy/mGy) or exposure-to-dose (mGy/R or rad/R) conversion factor, L is the length of the pencil ion chamber (i.e., 100 mm), and T is the slice thickness (mm). It can be shown that this method is mathematically equivalent to using a small exposure meter and acquiring all 14 CT scans. The 100-mm pencil ion chamber cannot be used to measure the CTDI_{FDA} for slice thicknesses other than 7 mm. However, it is commonly used for all slice thicknesses, and

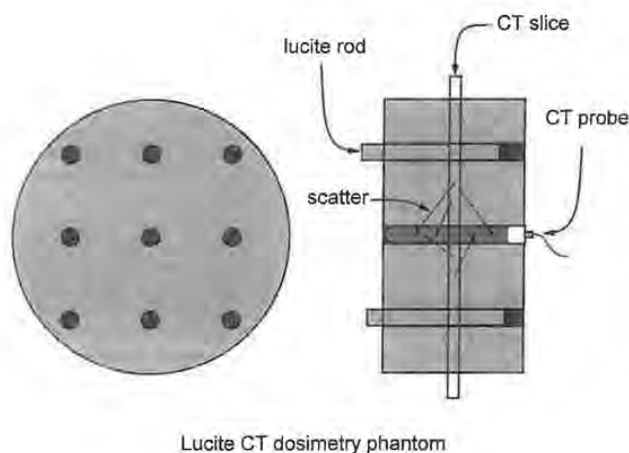


FIGURE 13-36. A diagram of a Lucite phantom, commonly used to determine dose in computed tomography (CT), is shown. The Lucite cylinder has holes drilled out for the placement of the CT pencil chamber. Lucite rods are used to plug in all remaining holes. A CT scan is acquired at the center of the cylinder, and dose measurements are made.

TABLE 13-1. TYPICAL COMPUTED TOMOGRAPHY DOSE INDEX DETERMINED DOSES AT 100 MAS TUBE CURRENT (AVERAGED FROM FOUR MANUFACTURERS DATA)

Peak Kilovoltage (kVp)	Head (16 cm diameter, mGy)	Body (32 cm diameter (mGy))
80	6.4	1.5
100	12.3	3.5
120	17.1	6.4
140	23.4	7.2

these measurements are referred to as the $CTDI_{100}$. The $CTDI_{FDA}$ significantly underestimates the MSAD for small slice thicknesses (e.g., 2 mm), because a significant amount of radiation is scattered beyond seven slice thicknesses. The $CTDI_{100}$ provides much better estimate of the MSAD for thin slices.

There are two standard phantoms that are commercially available for measuring the dose in CT. A 16-cm-diameter Lucite cylinder is used to simulate adult heads and pediatric torsos, and a 32-cm-diameter cylinder is used to simulate the adult abdomen (Fig. 13-36). Holes are drilled parallel to the long axis of the cylinder to allow the placement of the pencil chamber at different positions in the phantom. Lucite rods are used to plug the holes not occupied by the pencil chamber. The CTDI and MSAD are usually measured both at the center of each phantom and at 1 cm from the phantom's surface. The pencil ionization chamber measures air kerma, not dose directly, and the physicist converts this measurement to radiation dose (discussed in Chapter 3) specific to the x-ray beam energy and phantom material used. The CTDI and MSAD are typically expressed as dose to Lucite. For Lucite phantoms in typical CT beams, the conversion factor is approximately 0.893 mGy/mGy (air kerma), or 0.78 rad/roentgen. The dose to soft tissue is about 20% larger than the dose to Lucite. It is important to keep in mind that radiation dose in CT is proportional to the mAs used per slice. At the same kV, doubling of the mAs doubles the dose, and halving the mAs halves the dose. Some typical CTDIs are listed in Table 13-1.

Dose Considerations in Helical Scanning

Helical scanning with a collimator pitch of 1.0 is physically similar to performing a conventional (nonhelical) axial scan with contiguous slices. For CT scanners with multiple detector arrays, the collimator pitch (not the detector pitch) should be used for dose calculations. The dose in helical CT is calculated in exactly the same manner as it is with axial CT, using the CTDI discussed in the previous section; however a correction factor is needed when the pitch is not 1.0:

$$\text{Dose (helical)} = \text{Dose (axial)} \times \frac{1}{\text{Collimator pitch}}$$

For example, with a collimator pitch of 1.5, the dose from helical scanning is 67% that of the dose from conventional scanning, and a collimator pitch of 0.75 corresponds to a helical dose that is 133% of the axial dose (i.e., 33% greater), assuming the same mAs was used (Fig. 13-37). It is important to keep in mind, as mentioned earlier, that the dose changes linearly with the mAs of the study. Because

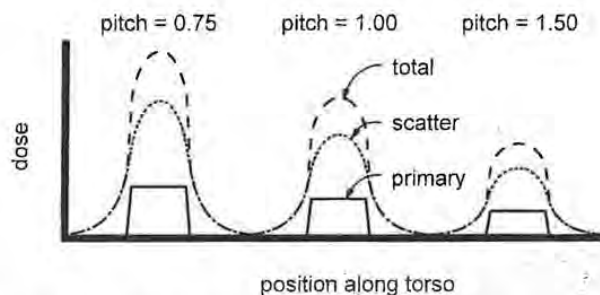


FIGURE 13-37. Dose profiles for three different collimator pitches are illustrated. A pitch of 1.0 corresponds to a dose profile very similar to that of an axial scan. For the same mAs and kV, a (collimator) pitch of 0.75 results in 33% more dose and a pitch of 1.5 corresponds to 66% less dose, compared with a pitch of 1.0. The primary, scatter, and total dose components are shown.

helical scans often use less mAs per 360-degree gantry rotation than axial scans do, the mAs should be factored into the calculation if it differs from the mAs used to measure the CTDI.

Dose in Computed Tomographic Fluoroscopy

Unlike most other CT studies, CT fluoroscopy involves repeated scanning of the same volume of tissue. In order to reduce the dose for these procedures, the mA used for CT fluoroscopic procedures is substantially reduced compared with that used for a regular CT scan. A CT fluoroscopic procedure using 20 mAs per gantry rotation results in a dose that is ten times less than that of a conventional CT scan taken at 200 mAs. The challenge to the radiologist is to perform the procedure, typically a biopsy, using as few seconds of CT fluoroscopy as possible. The dose calculation is straightforward:

$$\text{Dose (fluoro)} = \text{CTDI dose} \times \frac{\text{Time} \times \text{Current}}{\text{mAs of CTDI measurement}}$$

where time is given in seconds and the current (of the CT fluoroscopic examination) in mA. In the example, if 20 seconds of CT fluoroscopy is used at 20 mA and the CTDI (measured at the same kV) was 32 mGy for a 200-mAs study, then the dose would be $32 \text{ mGy} \times (20 \text{ sec} \times 20 \text{ mA}) / 200 \text{ mAs}$, or 64 mGy.

Current Modulation in Computed Tomography

Several CT manufacturers have introduced scanners that are capable of modulating the mA during the scan. In axial cross section, the typical human body is wider than it is thick. The mA modulation technique capitalizes on this fact. The rationale behind this technique is that it takes fewer x-ray photons (lower mA) to penetrate thinner tissue, and more x-ray photons (higher mA) are needed to penetrate thicker projections through the body. The SNR in the final image is related to the number of x-rays that pass through the patient and are detected. However, because of the way in which the filtered backprojection technique combines data from all views,

the high SNR detected in the thinner angular projections is essentially wasted, because the noise levels from the thicker angular projections dominate in the final image. Therefore, a reduction in patient dose can be achieved with little loss in image quality by reducing the mA during acquisition through the thinner tissue projections. The other benefit of this technique is that because the mA is reduced per gantry rotation, x-ray tube loading is reduced and helical scans can be performed for longer periods and with greater physical coverage. The mA modulation technique uses the angular-dependent signal levels (of the detectors) from the first few rotations of the gantry during acquisition (helical and axial) to modulate the mA during subsequent gantry rotations.

13.8 IMAGE QUALITY

Compared with x-ray radiography, CT has significantly worse spatial resolution and significantly better contrast resolution. The MTF, the fundamental measurement of spatial resolution, was shown in Fig. 13-30 for a typical CT scanner; it should be compared with the typical MTF for radiography, discussed in Chapter 10. Whereas the limiting spatial frequency for screen-film radiography is about 7 line pairs (lp) per millimeter and for digital radiography it is 5 lp/mm, the limiting spatial frequency for CT is approximately 1 lp/mm.

It is the contrast resolution of CT that distinguishes the modality: CT has, by far, the best contrast resolution of any clinical x-ray modality. Contrast resolution refers to the ability of an imaging procedure to reliably depict very subtle differences in contrast. It is generally accepted that the contrast resolution of screen-film radiography is approximately 5%, whereas CT demonstrates contrast resolution of about 0.5%. A classic clinical example in which the contrast resolution capability of CT excels is distinguishing subtle soft tissue tumors: The difference in CT number between the tumor and the surrounding tissue may be small (e.g., 20 CT numbers), but because the noise in the CT numbers is smaller (e.g., 3 CT numbers), the tumor is visible on the display to the trained human observer. As is apparent from this example, contrast resolution is fundamentally tied to the SNR. The SNR is also very much related to the number of x-ray quanta used per pixel in the image. If one attempts to reduce the pixel size (and thereby increase spatial resolution) and the dose levels are kept the same, the number of x-rays per pixel is reduced. For example, for the same FOV and dose, changing to a $1,024 \times 1,024$ CT image from a 512×512 image would result in fewer x-ray photons passing through each voxel, and therefore the SNR per pixel would drop. It should be clear from this example that there is a compromise between spatial resolution and contrast resolution. In CT there is a well-established relationship among SNR, pixel dimensions (Δ), slice thickness (T), and radiation dose (D):

$$D \propto \frac{SNR^2}{\Delta^3 T}$$

The clinical utility of any modality lies in its spatial and contrast resolution. In this chapter, many of the factors that affect spatial and contrast resolution of CT have been discussed. Below is a summary of the various factors affecting spatial and contrast resolution in CT.

Factors Affecting Spatial Resolution

Detector pitch: The detector pitch is the center-to-center spacing of the detectors along the array. For third-generation scanners, the detector pitch determines the ray spacing; for fourth-generation scanners, the detector pitch influences the view sampling.

Detector aperture: The detector aperture is the width of the active element of one detector. The use of smaller detectors increases the cutoff (Nyquist) frequency of the image, and it improves spatial resolution at all frequencies.

Number of views: The number of views influences the ability of the CT image to convey the higher spatial frequencies in the image without artifacts (see Fig. 13-23). Use of too few views results in view aliasing, which is most noticeable toward the periphery of the image.

Number of rays: The number of rays used to produce a CT image over the same FOV has a strong influence on spatial resolution (see Fig. 13-22). For a fixed FOV, the number of rays increases as the detector pitch decreases.

Focal spot size: As in any x-ray imaging procedure, larger focal spots cause more geometric unsharpness in the detected image and reduce spatial resolution. The influence of focal spot size is very much related to the magnification of an object to be resolved.

Object magnification: Increased magnification amplifies the blurring of the focal spot. Because of the need to scan completely around the patient in a fixed-diameter gantry, the magnification factors experienced in CT are higher than in radiography. Magnification factors of 2.0 are common, and they can reach 2.7 for the entrant surfaces of large patients.

Slice thickness: The slice thickness is equivalent to the detector aperture in the cranial-caudal axis. Large slice thicknesses clearly reduce spatial resolution in the cranial-caudal axis, but they also reduce sharpness of the edges of structures in the transaxial image. If a linear, high-contrast object such as a contrast-filled vessel runs perfectly perpendicular to the transaxial plane, it will exhibit sharp edges regardless of slice thickness. However, if it traverses through the patient at an angle, its edges will be increasingly blurred with increased slice thickness.

Slice sensitivity profile: The slice sensitivity profile is quite literally the line spread function in the cranial-caudal axis of the patient. The slice sensitivity profile is a more accurate descriptor of the slice thickness.

Helical pitch: The pitch used in helical CT scanning affects spatial resolution, with greater pitches reducing resolution. A larger pitch increases the width of the slice sensitivity profile.

Reconstruction kernel: The shape of the reconstruction kernel has a direct bearing on spatial resolution. Bone filters have the best spatial resolution, and soft tissue filters have lower spatial resolution.

Pixel matrix: The number of pixels used to reconstruct the CT image has a direct influence (for a fixed FOV) on spatial resolution; however no CT manufacturer compromises on this parameter in a way that reduces resolution. In some instances the pixel matrix may be reduced. For example, some vendors reconstruct to a 256×256 pixel matrix for CT fluoroscopy to achieve real-time performance. Also, in soft copy viewing, if the number of CT images displayed on a monitor exceeds the pixel resolution of the display hardware, the software downscales the CT images, reducing the spatial resolution. For example, a $1,024 \times 1,024$ pixel workstation can display only four 512×512 CT images at full resolution.

Patient motion: If there is involuntary motion (e.g., heart) or motion resulting from patient noncompliance (e.g., respiratory), the CT image will experience blurring proportional to the distance of the motion during the scan.

Field of view: The FOV influences the physical dimensions of each pixel. A 10-cm FOV in a 512×512 matrix results in pixel dimensions of approximately 0.2 mm, and a 35-cm FOV produces pixel widths of about 0.7 mm.

Factors Affecting Contrast Resolution

mAs: The mAs directly influence the number of x-ray photons used to produce the CT image, thereby affecting the SNR and the contrast resolution. Doubling of the mAs of the study increases the SNR by $\sqrt{2}$ or 41%, and the contrast resolution consequently improves.

Dose: The dose increases linearly with mAs per scan, and the comments for mAs apply here.

Pixel size (FOV): If patient size and all other scan parameters are fixed, as FOV increases, pixel dimensions increase, and the number of x-rays passing through each pixel increases.

Slice thickness: The slice thickness has a strong (linear) influence on the number of photons used to produce the image. Thicker slices use more photons and have better SNR. For example, doubling of the slice thickness doubles the number of photons used (at the same kV and mAs), and increases the SNR by $\sqrt{2}$, or 41%.

Reconstruction filter: Bone filters produce lower contrast resolution, and soft tissue filters improve contrast resolution.

Patient size: For the same x-ray technique, larger patients attenuate more x-rays, resulting in detection of fewer x-rays. This reduces the SNR and therefore the contrast resolution.

Gantry rotation speed: Most CT systems have an upper limit on mA, and for a fixed pitch and a fixed mA, faster gantry rotations (e.g., 1 second compared with 0.5 second) result in reduced mAs used to produce each CT image, reducing contrast resolution.

13.9 ARTIFACTS

Beam Hardening

Like all medical x-ray beams, CT uses a polyenergetic x-ray spectrum, with energies ranging from about 25 to 120 keV. Furthermore, x-ray attenuation coefficients are energy dependent: After passing through a given thickness of patient, lower-energy x-rays are attenuated to a greater extent than higher-energy x-rays are. Therefore, as the x-ray beam propagates through a thickness of tissue and bone, the shape of the spectrum becomes skewed toward the higher energies (Fig. 13-38). Consequently, the average energy of the x-ray beam becomes greater ("harder") as it passes through tissue. Because the attenuation of bone is greater than that of soft tissue, bone causes more beam hardening than an equivalent thickness of soft tissue (Fig. 13-39).

The beam-hardening phenomenon induces artifacts in CT because rays from some projection angles are hardened to a differing extent than rays from other angles, and this confuses the reconstruction algorithm. A calculated example is shown in Fig. 13-40A. The most common clinical example of beam hardening

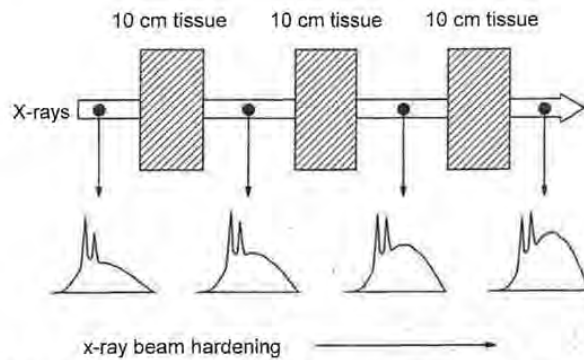


FIGURE 13-38. The nature of x-ray beam hardening is illustrated. As a spectrum of x-rays (**lower graphs**) passes layers of tissue, the lower-energy photons in the x-ray spectrum are attenuated to a greater degree than the higher-energy components of the spectrum. Therefore, as the spectrum passes through increasing thickness of tissue, it becomes progressively skewed toward the higher-energy x-rays in that spectrum. In the vernacular of x-ray physics, a higher-energy spectrum is called a "harder" spectrum; hence the term *beam hardening*.

occurs between the petrous bones in the head, where a spider web-like artifact connects the two bones on the image.

Most CT scanners include a simple beam-hardening correction algorithm that corrects for beam hardening based on the relative attenuation of each ray, and this helps to some extent. More sophisticated two-pass beam-hardening correction algorithms have been implemented by some manufacturers, and these correction techniques can substantially reduce artifacts resulting from beam hardening. In a two-pass algorithm, the image is reconstructed normally in the first pass, the path length

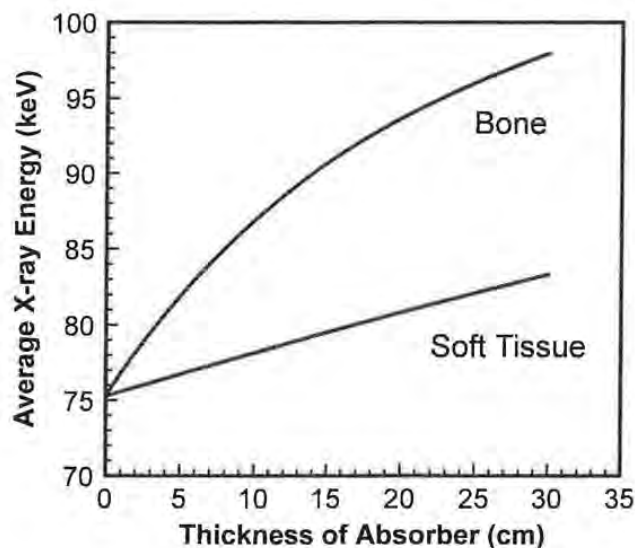


FIGURE 13-39. Using a 120-kV x-ray spectrum filtered by 7 mm of aluminum as the incident x-ray spectrum, the average x-ray energy of the spectrum is shown as a function of the thickness of the absorber through which the spectrum propagates. Beam hardening caused by soft tissue and by bone is illustrated. Bone causes greater beam hardening because of its higher attenuation coefficient (due to increased density and higher average atomic number).

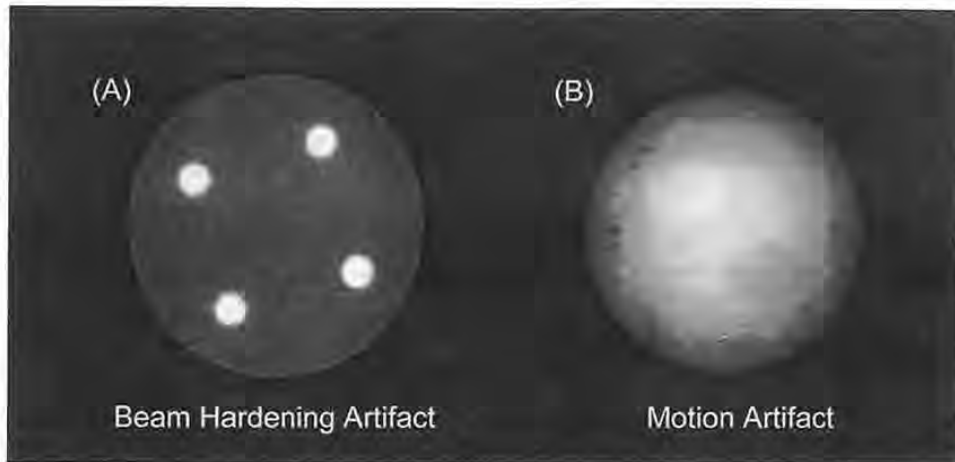


FIGURE 13-40. A: A beam-hardening artifact caused by four calcium-filled cylinders is illustrated. **B:** Artifacts due to motion are shown.

that each ray transits through bone and soft tissue is determined from the first-pass image, and then each ray is compensated for beam hardening based on the known bone and soft tissue thicknesses it has traversed. A second-pass image reconstruction is performed using the corrected ray values.

Motion Artifacts

Motion artifacts occur when the patient moves during the acquisition. Small motions cause image blurring, and larger physical displacements during CT image acquisition produce artifacts that appear as double images or image ghosting. If motion is suspected in a CT image, the adjacent CT scans in the study may be evaluated to distinguish fact from artifact. In some cases, the patient needs to be re-scanned. An example of a motion artifact is shown in Fig. 13-40B.

Partial Volume Averaging

The CT number in each pixel is proportional to the average μ in the corresponding voxel. For voxels containing all one tissue type (e.g., all bone, all liver), μ is representative of that tissue. Some voxels in the image, however, contain a mixture of different tissue types. When this occurs, for example with bone and soft tissue, the μ is not representative of either tissue but instead is a weighted average of the two different μ values. Partial volume averaging is most pronounced for softly rounded structures that are almost parallel to the CT slice. The most evident example is near the top of the head, where the cranium shares a substantial number of voxels with brain tissue, causing details of the brain parenchyma to be lost because the large μ of bone dominates. This situation is easily recognizable and therefore seldom leads to misdiagnosis. Partial volume artifacts can lead to misdiagnosis when the presence of adjacent anatomic structures is not suspected (Fig. 13-41).

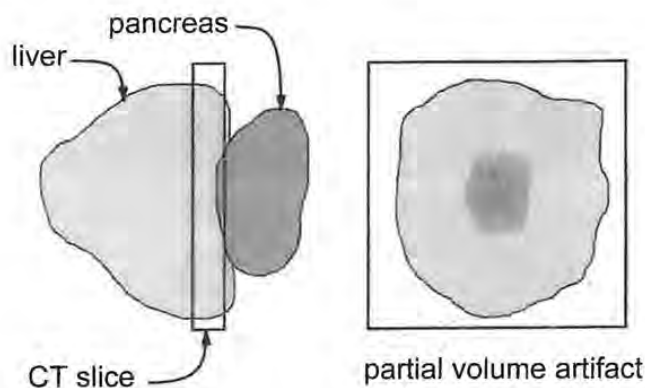


FIGURE 13-41. A partial volume artifact occurs when the computed tomographic slice interrogates a slab of tissue containing two or more different tissue types. Many partial volume artifacts are obvious (e.g., with bone), but occasionally a partial volume artifact can mimic pathologic conditions.

There are several approaches to reducing partial volume artifacts, and the obvious *a priori* approach is to use thinner CT slices. When a suspected partial volume artifact occurs with a helical study and the raw scan data is still available, it is sometimes necessary to use the raw data to reconstruct additional CT images at different positions. Many institutions make use of interleaved reconstructions (e.g., 5-mm slices every 2.5 mm) as a matter of routine protocol.

SUGGESTED READING

- Cho ZH, Jones JP, Singh M. *Foundations of medical imaging*. New York: John Wiley and Sons, 1993.
- Goldman LW, Fowlkes JB. *Syllabus from the 2000 categorical course in diagnostic radiology physics: CT and US cross sectional imaging*. Oak Brook, IL: RSNA Publications, 2000.
- Herman GT. *Image reconstruction from projections: the fundamentals of computerized tomography*. New York: Academic Press, 1980.
- Huda W, Atherton JV. Energy imparted in computed tomography. *Med Phys* 1994;22:1263–1269.
- Sprawls P. AAPM tutorial: CT image detail and noise. *Radiographics* 1992;12:1041–1046.

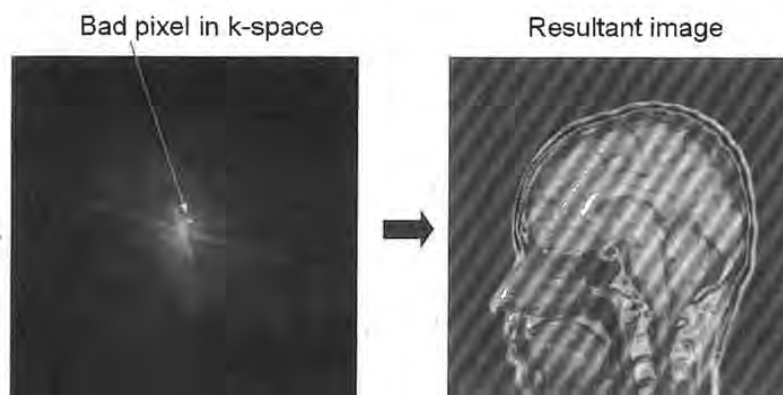


FIGURE 15-33. A single bad pixel in the k-space causes a significant artifact in the reconstructed image. The bad pixel is located at $k_x = 2$, $k_y = 3$, which produces a superimposed sinusoidal wave on the spatial domain image.

Motion Artifacts

The most ubiquitous and noticeable artifacts in MRI arise with patient motion, including voluntary and involuntary movement, and flow (blood, CSF). Although motion artifacts are not unique to MRI, the long acquisition time of certain MRI sequences increases the probability of motion blurring and contrast resolution losses. Motion artifacts mostly occur along the phase encode direction, since adjacent lines of phase-encoded protons are separated by a TR interval that can last 3,000 msec or longer. Even very slight motion will cause a change in the recorded phase variation across the FOV throughout the MR acquisition sequence (Fig. 15-34). The frequency encode direction is less affected, especially by periodic motion,



FIGURE 15-34. Motion artifacts, particularly of flow patterns, are most always displayed in the phase encode gradient direction. Slight changes in phase produce multiple ghost images of the anatomy, since the variation in phase caused by motion can be substantial between excitations.

since the evolution of the echo signal, frequency encoding, and sampling occur simultaneously over several milliseconds as opposed to seconds to minutes in the phase direction. Ghost images, which are faint copies of the image displaced along the phase encode direction, are the visual result of patient motion.

Several techniques can compensate for motion-related artifacts. The simplest technique transposes the PEG and FEG to relocate the motion artifacts out of the region of diagnostic interest with the same pulse sequences. This does not reduce the magnitude of the artifacts, however.

There are other motion compensation methods:

1. Cardiac and respiratory gating—signal acquisition at a particular cyclic location synchronizes the phase changes applied across the anatomy (Fig. 15-35).
2. Respiratory ordering of the phase encoding projections based on location within the respiratory cycle.
3. Signal averaging to reduce artifacts of random motion by making displaced signals less conspicuous relative to stationary anatomy.
4. Short TE spin echo sequences (limited to proton density, T1-weighted scans). Long TE scans (T2 weighting) are more susceptible to motion.
5. Gradient moment nulling (additional gradient pulses for flow compensation) to help rephase spins that are dephased due to motion. Most often, these techniques require a longer TE, and are more useful for T2-weighted scans (Fig. 15-25).
6. Presaturation pulses applied outside the imaging region to reduce flow artifacts from inflowing spins (Fig. 15-28).

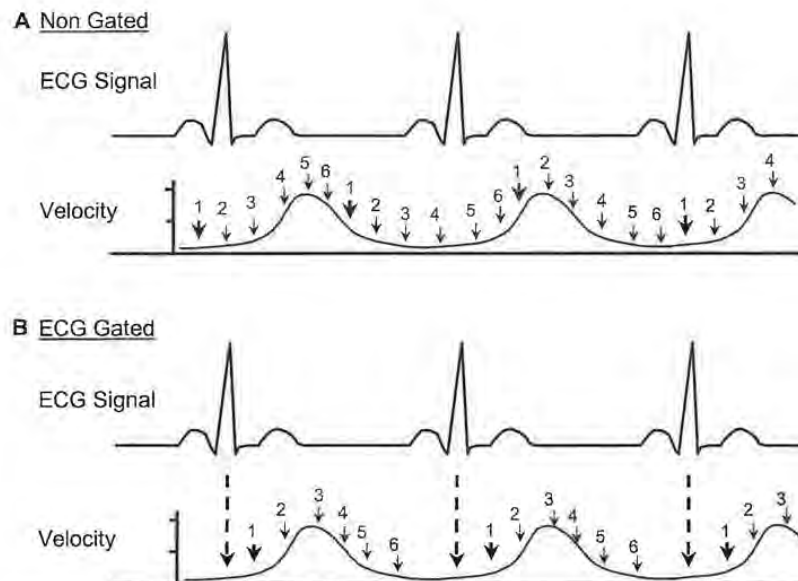


FIGURE 15-35. A: Motion artifacts occur because the multislice image data (indicated as 1 through 6) are not synchronized with the periodicity of the beating heart and the velocity pattern. **B:** Electrocardiogram gated acquisitions synchronize the acquisition of the image data with the R wave (peak wave of ECG signal) to reduce vascular motion artifacts; a reduced number of images or extended acquisition time is required.

On T1-weighted images of the head, blood flow and CSF pulsations are the source of most motion artifacts. Pulse sequences using a gradient moment nulling technique "refocus" moving spins by using switched readout gradient polarities, and these techniques can correct for periodic motion occurring during the TE period. Similarly, on T2-weighted head images the most detrimental motion degradations arise from pulsating CSF. Gradient moment nulling techniques can also be applied to null the signal from these moving spins and improve T2 contrast. Compensation can be applied to one, two, or three axes, and corrections can be first order (linear velocity), second order (acceleration), or higher order, depending on the applied readout gradient waveform. ECG pulse and respiratory gating is a complementary technique that significantly reduces cardiac, CSF, and respiratory motion. The trade-offs of gating are increased image acquisition and operator setup time.

Chemical Shift Artifacts

Chemical shift refers to the resonance frequency variations resulting from intrinsic magnetic shielding of anatomic structures. Molecular structure and electron orbital characteristics produce fields that shield the main magnetic field and give rise to distinct peaks in the MR spectrum. In the case of proton spectra, peaks correspond to water and fat, and in the case of breast imaging, silicone material. Lower frequencies of about 3.5 parts per million for protons in fat and 5.0 parts per million for protons in silicone occur, compared to the resonance frequency of protons in water (Fig. 15-36A). Since resonance frequency increases linearly with field strength, the absolute difference between the fat and water resonance also increases, making high field strength magnets more susceptible to chemical shift artifact. A sample calculation in the example below demonstrates frequency variations in fat and water for two different magnetic field and gradient field strengths.

Data acquisition methods cannot directly discriminate a frequency shift due to the application of a frequency encode gradient or to a chemical shift artifact. Water and fat differences therefore cannot be distinguished by the frequency difference induced by the gradient. The protons in fat resonate at a slightly lower frequency than the corresponding protons in water, and cause a shift in the anatomy (misregistration of water and fat moieties) along to the frequency encode gradient direction (Fig. 15-36B,C).

Example: Numerical calculation of the chemical shift artifact

Field strength:

A 3 parts per million (ppm) chemical shift of water and fat is equivalent to a frequency difference at:

0.5 T of $21 \text{ MHz} \times 3 \text{ ppm} = 63 \text{ Hz}$

1.5 T of $64 \text{ MHz} \times 3 \text{ ppm} = 192 \text{ Hz}$

Thus, the chemical shift is more severe for higher field strength magnets.

Gradient strength, 25 cm (0.25 m) FOV, 256×256 matrix:

A gradient strength of $2.5 \text{ mT/m} \times 0.25 \text{ m} = 0.000625 \text{ T}$

Frequency difference: $0.000625 \text{ T} \times 42.58 \text{ MHz/T} = 26.6 \text{ kHz/256 pixels}$
 $= 104 \text{ Hz/pixel}$

A gradient strength of $10 \text{ mT/m} \times 0.25 \text{ m} = 0.0025 \text{ T}$

Frequency difference: $0.0025 \text{ T} \times 42.58 \text{ MHz/T} = 106.5 \text{ kHz/256 pixels}$
 $= 416 \text{ Hz/pixel}$

Thus, a chemical shift is more severe for lower gradient strengths.

ufacturers incorporate a computer for image acquisition and camera control in the camera itself and provide a separate computer for image processing and display, whereas others provide a single computer for both purposes.

Some older scintillation cameras transmit pairs of analog X- and Y-position pulses to the computer. A computer interfaced to such a camera has an acquisition interface containing two ADCs for converting the pairs of analog position pulses into digital form. However, in most modern scintillation cameras, the pairs of X and Y signals are created using digital circuitry and are transferred from the camera to the computer in digital form. The formation of a digital image is described in the following section.

Digital Image Formats in Nuclear Medicine

As discussed in Chapter 4, a digital image consists of a rectangular array of numbers, and an element of the image represented by a single number is called a *pixel*. In nuclear medicine, each pixel represents the number of counts detected from activity in a specific location within the patient. Common image formats in nuclear medicine are 64^2 and 128^2 pixels. Whole-body images are stored in larger formats, such as 256 by 1,024 pixels. If one byte (8 bits) is used for each pixel, the image is said to be in *byte mode*; if two bytes are used for each pixel, the image is said to be in *word mode*. A single pixel can store as many as 255 counts in byte mode and 65,535 counts in word mode. Modern nuclear medicine computers may allow only word-mode images.

Image Acquisition

Frame Mode Acquisition

Image data in nuclear medicine are acquired in either frame or list mode. Figure 21-21 illustrates frame-mode acquisition. Before acquisition begins, a portion of the computer's memory is designated to contain the image. All pixels within this image are set to zero. After acquisition begins, pairs of X- and Y-position signals are received from the camera. Each pair of numbers designates a single pixel in the image. One count is added to the counts in that pixel. As many pairs of position signals are received, the image is formed.

There are three types of frame-mode acquisition: *static*, *dynamic*, and *gated*. In a static acquisition, a single image is acquired for either a preset time interval or until the total number of counts in the image reaches a preset number. In a dynamic acquisition, a series of images is acquired one after another, for a preset time per image. Dynamic image acquisition is used to study dynamic processes, such as the first transit of a bolus of a radiopharmaceutical through the heart or the extraction and excretion of a radiopharmaceutical by the kidneys.

Some dynamic processes occur too rapidly for them to be effectively portrayed by dynamic image acquisition; each image in the sequence would have too few counts to be statistically valid. However, if the dynamic process is repetitive, gated image acquisition may permit the acquisition of an image sequence that accurately depicts the dynamic process. Gated acquisition is most commonly used for evalu-

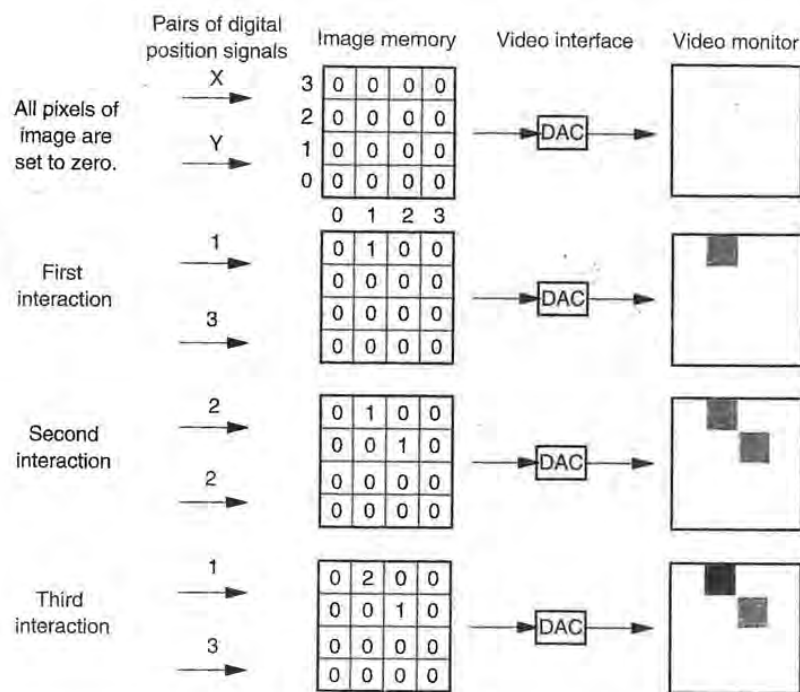


FIGURE 21-21. Acquisition of an image in frame mode. This figure illustrates the incorporation of position data from the first three interactions into an image.

ating cardiac mechanical performance, in cardiac blood pool studies and, sometimes, in myocardial perfusion studies. Gated acquisition requires a physiologic monitor that provides a trigger pulse to the computer at the beginning of each cycle of the process being studied. In gated cardiac studies, an electrocardiogram (ECG) monitor provides a trigger pulse to the computer whenever the monitor detects a QRS complex.

Figure 21-22 depicts the acquisition of a gated cardiac image sequence. First, space for the desired number of images (usually 16 to 24) is reserved in the computer's memory. Next, several cardiac cycles are timed and the average time per cycle is determined. The time per cycle is divided by the number of images in the sequence to obtain the time T per image. Then the acquisition begins. When the first trigger pulse is received, the acquisition interface sends the data to the first image in the sequence for a time T . Then it is stored in the second image in the sequence for a time T , after which it is stored in the third image for a time T . This process proceeds until the next trigger pulse is received. Then the process begins anew, with the data being added to that in the first image for a time T , then the second image for a time T , etc. This process is continued until a preset time interval, typically 10 minutes, has elapsed, enabling sufficient counts to be collected for the image sequence to form a statistically valid depiction of an average cardiac cycle.

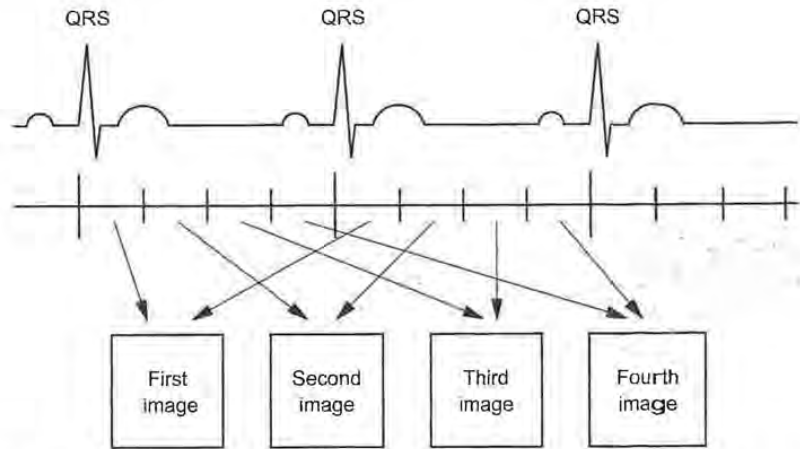


FIGURE 21-22. Acquisition of a gated cardiac image sequence. Only four images are shown here. Sixteen to 24 images are typically acquired.

List-Mode Acquisition

In *list-mode acquisition*, the pairs of X- and Y-position values are stored in a list (Fig. 21-23, instead of being immediately formed into an image. Periodic timing marks are included in the list. If a physiologic monitor is being used, as in gated cardiac imaging, trigger marks are also included in the list. After acquisition is complete, the list-mode data are reformatted into conventional images for display. The advantage of list-mode acquisition is that it allows great flexibility in how the X and Y values are combined to form images. The disadvantages of list-mode acquisition are that it generates large amounts of data, requiring more memory to acquire and disk space to store than frame-mode images, and that the data must be subsequently processed into standard images for viewing.

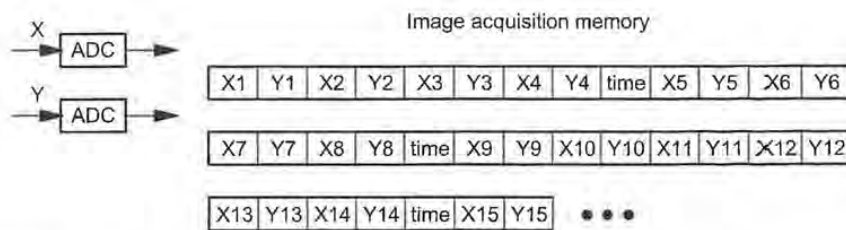


FIGURE 21-23. Acquisition of image data in list mode. The digital X- and Y-position signals are stored as a list. Periodically, a timing signal is also stored.

Image Processing in Nuclear Medicine

A major reason for the use of computers in nuclear medicine is that they provide the ability to present the data in the unprocessed images in ways that are of greater use to the clinician. Although it is not within the scope of this section to provide a comprehensive survey of image processing in nuclear medicine, the following are common examples.

Image Subtraction

When one image is subtracted from another, each pixel count in one image is subtracted from the corresponding pixel count in the other. Negative numbers resulting from these subtractions are set to zero. The resultant image depicts the change in activity that occurs in the patient during the time interval between the acquisitions of the two images.

Regions of Interest and Time-Activity Curves

A *region of interest* (ROI) is a closed boundary that is superimposed on the image. It may be drawn manually or it may be drawn automatically by the computer. The sum of the counts in all pixels in the ROI is an indication of the activity in the corresponding portion of the patient.

To create a *time-activity curve* (TAC), a ROI must first be drawn on one image of a dynamic or gated image sequence. The same ROI is then superimposed on each image in the sequence by the computer and the total number of counts within the ROI is determined for each image. Finally, the counts within the ROI are plotted as a function of image number. The resultant curve is an indication of the activity in the corresponding portion of the patient as a function of time. Figure 21-24 shows ROIs over both kidneys of a patient and time-activity curves describing the uptake and excretion of the radiopharmaceutical Tc-99m MAG-3 by the kidneys.

Spatial Filtering

Nuclear medicine images have a grainy appearance because of the statistical nature of the acquisition process. This quantum mottle can be reduced by a type of spatial filtering called *smoothing*. Unfortunately, smoothing also reduces the spatial resolution of the image. Images should not be smoothed to the extent that clinically significant detail is lost. Spatial filtering is described in detail in Chapter 11.

Left Ventricular Ejection Fraction

The left ventricular ejection fraction (LVEF) is a measure of the mechanical performance of the left ventricle of the heart. It is defined as the fraction of the end-diastolic volume ejected during a cardiac cycle:

$$\text{LVEF} = (V_{\text{ED}} - V_{\text{ES}}) / V_{\text{ED}} \quad [21-10]$$

where V_{ED} is the end-diastolic volume and V_{ES} is the end-systolic volume of the ventricle. In nuclear medicine, it is usually determined from an equilibrium gated blood-

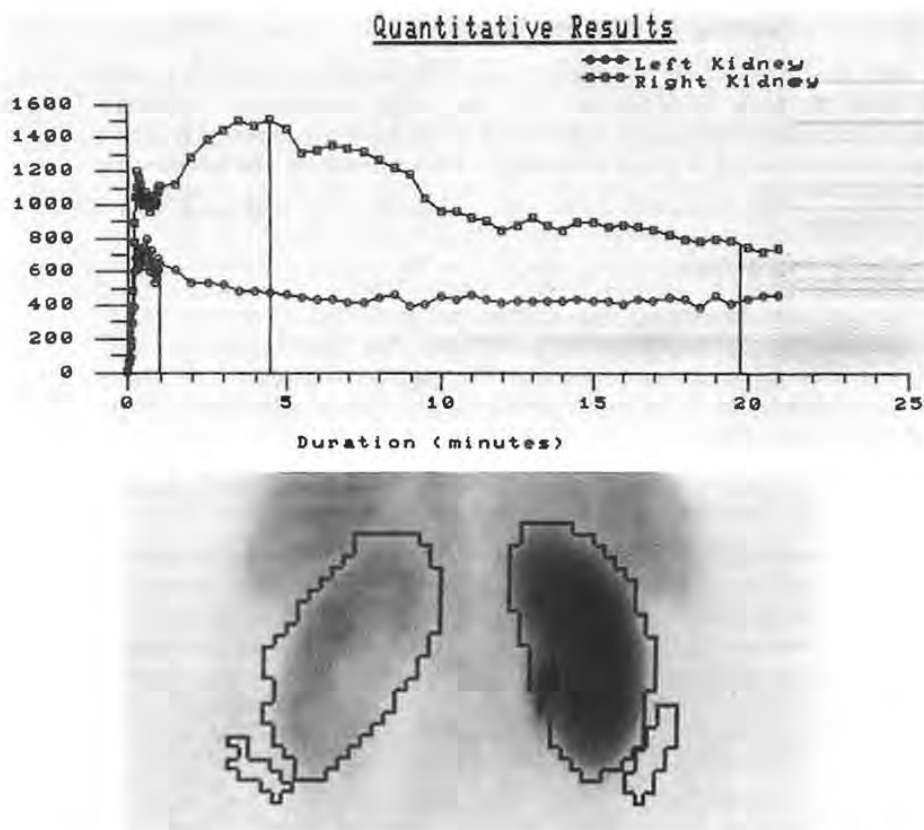


FIGURE 21-24. Regions of interest (ROIs) (bottom) and time-activity curves (TACs) (top).

pool image sequence, using Tc-99m-labeled red blood cells. The image sequence is acquired from a left anterior oblique (LAO) projection, with the camera positioned at the angle demonstrating the best separation of the two ventricles, after sufficient time has elapsed for the administered activity to reach a uniform concentration in the blood.

The calculation of the LVEF is based on the assumption that the counts from left ventricular activity are approximately proportional to the ventricular volume throughout the cardiac cycle. A ROI is first drawn around the left ventricular cavity, and a TAC is obtained by superimposing this ROI over all images in the sequence. The first image in the sequence depicts end diastole, and the image containing the least counts in the ROI depicts end systole. The total left ventricular counts in the end-diastolic and end-systolic images are determined. Some programs use the same ROI around the left ventricle for both images, whereas, in other programs, the ROI is drawn separately in each image to better fit the varying shape of the ventricle. Unfortunately, each of these counts is due not only to activity in the left ventricular cavity but also to activity in surrounding tissues, chambers, and great vessels. To compensate for this "crosstalk" (commonly called "background activity"), another ROI is drawn just beyond the wall of the left ventricle, avoiding active

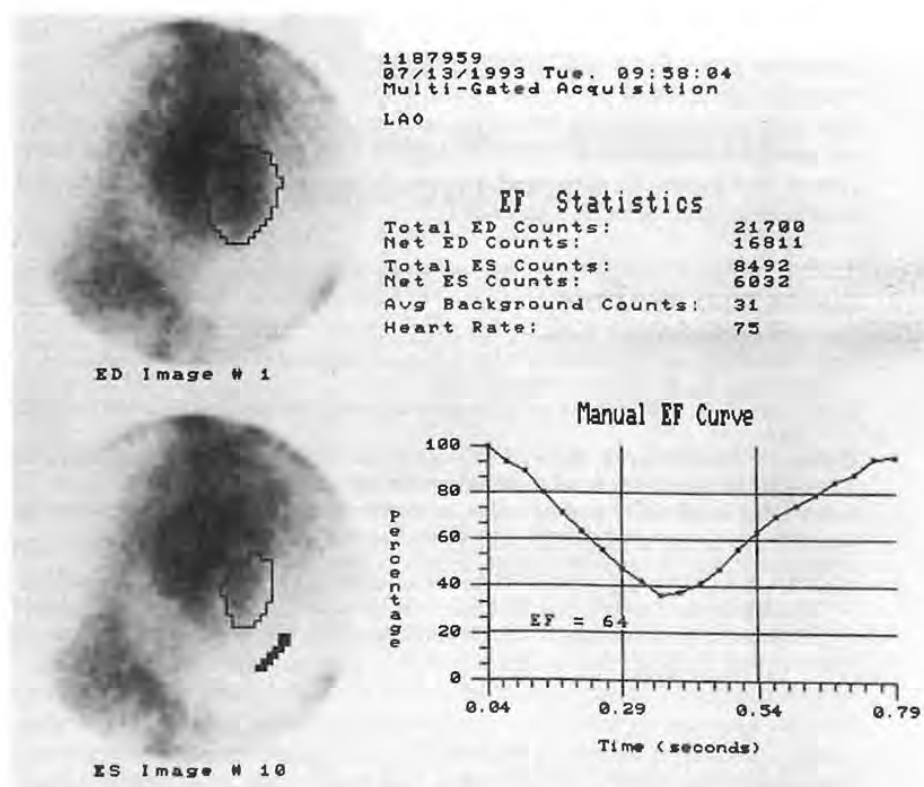


FIGURE 21-25. End-diastolic (**top**) and end-systolic (**bottom**) images from a gated cardiac blood-pool study, showing ROIs used to determine the left ventricular ejection fraction (LVEF). The small dark ROI below and to the right of the LV ROI in the end-systolic image is used to determine and correct for the extracardiac ("background") counts.

structures such as the spleen, cardiac chambers, and great vessels. The number of counts in the left-ventricular ROI due to crosstalk is estimated as follows:

$$\text{Counts crosstalk} = \frac{(\text{Counts in crosstalk ROI}) (\text{Pixels in LV ROI})}{(\text{Pixels in crosstalk ROI})} \quad [21-11]$$

Figure 21-25 shows end-diastolic and end-systolic images of the heart from the LAO projection and the ROIs used to determine the LVEF. The LVEF is then calculated using the following equation:

$$\text{LVEF} = \frac{(\text{Counts ED}) - (\text{Counts ES})}{(\text{Counts ED}) - (\text{Counts crosstalk})} \quad [21-12]$$

SUMMARY

This chapter described the principles of operation of the Anger scintillation camera, which forms images, depicting the distribution of x- and gamma-ray-emitting

radionuclides in patients, using a planar crystal of the scintillator NaI(Tl) optically coupled to a two-dimensional array of photomultiplier tubes. The collimator, necessary for formation of projection images, imposes a compromise between spatial resolution and sensitivity in detecting x- and gamma rays. Because of the collimator, the spatial resolution of the images degrades with distance from the face of the camera. In Chapter 22, the use of the scintillation camera to perform computed tomography, called *SPECT*, is described.

SUGGESTED READING

- Anger HO. Radioisotope cameras. In: Hine GJ, ed. *Instrumentation in nuclear medicine*, vol. 1. New York: Academic Press, 1967:485–552.
- Gelfand MJ, Thomas SR. *Effective use of computers in nuclear medicine*. New York: McGraw-Hill, 1988.
- Graham LS, Muehllehner G. Anger scintillation camera. In: Sandler MP, et al., eds. *Diagnostic nuclear medicine*, 3rd ed., vol. 1. Baltimore: Williams & Wilkins, 1996:81–92.
- Groch MW. Cardiac function: gated cardiac blood pool and first pass imaging. In: Henkin RE, et al., eds. *Nuclear medicine*, vol. 1. St. Louis: Mosby, 1996:626–643.
- Simmons GH, ed. *The scintillation camera*. New York: Society of Nuclear Medicine, 1988.
- Yester MV, Graham LS, eds. *Advances in nuclear medicine: the medical physicist's perspective*. Proceedings of the 1998 Nuclear Medicine Mini Summer School, American Association of Physicists in Medicine, June 21–23, 1998, Madison, WI.

NUCLEAR IMAGING—EMISSION TOMOGRAPHY

The formation of projection images in nuclear medicine was discussed in Chapter 21. A nuclear medicine projection image depicts a two-dimensional projection of the three-dimensional activity distribution in the patient. The disadvantage of a projection image is that the contributions to the image from structures at different depths overlap, hindering the ability to discern the image of a structure at a particular depth. Tomographic imaging is fundamentally different; it attempts to depict the activity distribution in a single cross section of the patient.

There are two fundamentally different types of tomography: conventional tomography, also called geometric or focal plane tomography, and computed tomography. In conventional tomography, structures out of a focal plane are not removed from the resultant image; instead, they are blurred by an amount proportional to their distance from the focal plane. Those close to the focal plane suffer little blurring and remain apparent in the image. Even those farther away, although significantly blurred, contribute to the image, thereby reducing contrast. In distinction, computed tomography (CT) uses mathematical methods to remove overlying structures completely. CT requires the acquisition of a set of projection images from at least a 180-degree arc about the patient. The projection images are then mathematically reconstructed by a computer to form images depicting cross sections of the patient. Just as in x-ray transmission imaging, both conventional and computed tomography are possible in nuclear medicine imaging. Single photon emission computed tomography (SPECT) and positron emission tomography (PET) are both forms of CT.

Focal Plane Tomography in Nuclear Medicine

Focal plane tomography once had a significant role in nuclear medicine but is seldom used today. The rectilinear scanner, when used with focused collimators, was an example of conventional tomography. A number of other devices have been developed to exploit conventional tomography in nuclear medicine. The Anger tomoscanner used two small scintillation cameras with converging collimators, one above and one below the patient table, to scan the patient in a raster pattern and produce multiple whole-body images, each showing structures at a different depth in the patient in focus, in a single scan. The seven-pinhole collimator was used with a conventional scintillation camera and computer to produce short-axis images of the heart, each showing structures at a different depth in focus. The rectilinear scanner and the Anger tomoscanner are no longer produced. The seven-pinhole colli-

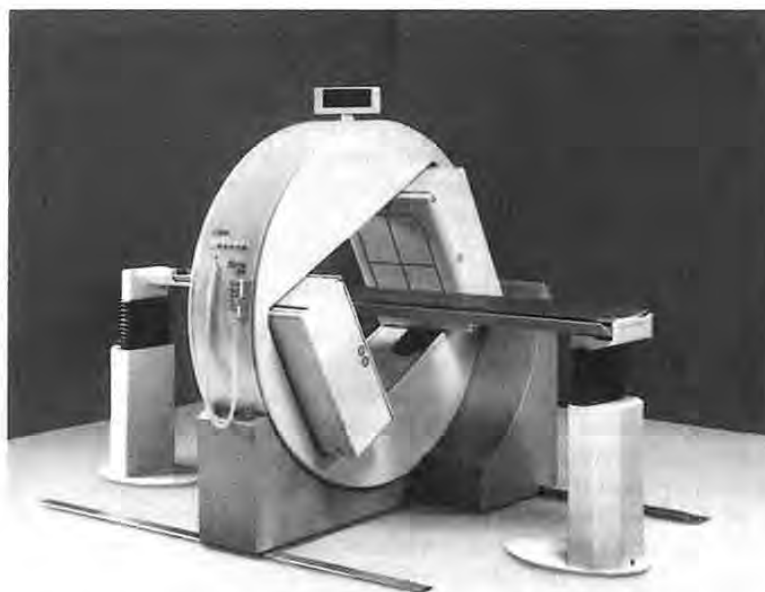
mator, which never enjoyed wide acceptance, has been almost entirely displaced by SPECT.

The scintillation camera itself, when used for planar imaging with a parallel-hole collimator, produces a weak tomographic effect. The system spatial resolution decreases with distance, causing structures farther from the camera to be more blurred than closer structures. This effect is perhaps most clearly evident in planar skeletal imaging of the body. In the anterior images, for example, the sternum and anterior portions of the ribs are clearly shown, while the spine and posterior ribs are barely evident. (Attenuation also plays a role in this effect.)

22.1 SINGLE PHOTON EMISSION COMPUTED TOMOGRAPHY (SPECT)

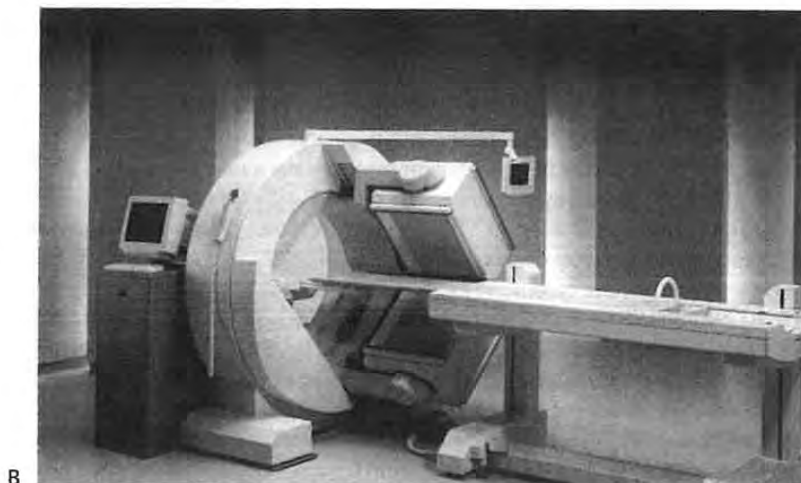
Design and Principles of Operation

Single photon emission computed tomography (SPECT) generates transverse images depicting the distribution of x- or gamma-ray emitting nuclides in patients. Standard planar projection images are acquired from an arc of 180 degrees (most cardiac SPECT) or 360 degrees (most noncardiac SPECT) about the patient. Although these images could be obtained by any collimated imaging device, the vast majority of SPECT systems use one or more scintillation camera heads that revolve about the patient. The SPECT system's digital computer then reconstructs the transverse images, using either filtered backprojection, as does the computer in an x-ray CT system, or iterative reconstruction methods, which are described later in this chapter. Figure 22-1 shows multihead SPECT systems. Figure 21-1 in Chapter 21 shows a single-head SPECT system.



A

FIGURE 22-1. A: Double-head, fixed 180-degree (photograph courtesy of Marconi Medical Systems).



B



C

FIGURE 22-1 (continued). (B) double-head, variable-angle (photograph courtesy of Siemens Medical Systems); (C) and triple-head (photograph courtesy of Tri-onix Research Laboratory, Inc.) single photon emission computed tomography (SPECT) systems. The double-head, variable-angle system is shown with the heads at a 76-degree angle for cardiac acquisition.

Image Acquisition

The camera head or heads of a SPECT system revolve about the patient, acquiring projection images from evenly spaced angles. The head or heads may acquire the images while moving (continuous acquisition) or may stop at predefined angles to acquire the images ("step and shoot" acquisition). Each projection image is acquired in a computer in frame mode (see Image Acquisition in Chapter 21). If the camera

heads of a SPECT system produced ideal projection images (i.e., no attenuation by the patient and no degradation of spatial resolution with distance from the camera), projection images from opposite sides of the patient would be mirror images and projection images over a 180-degree arc would be sufficient for transverse image reconstruction. However, in SPECT, attenuation greatly reduces the number of photons from activity in the half of the patient opposite the camera head, and this information is greatly blurred by the distance from the collimator. Therefore, for most noncardiac studies, such as brain SPECT, the projection images are acquired over a complete revolution (360 degrees) about the patient. However, most nuclear medicine laboratories acquire cardiac SPECT studies, such as myocardial perfusion studies, over a 180-degree arc from the 45-degree right anterior oblique (RAO) view to the 45-degree left posterior oblique (LPO) view (Fig. 22-2). The 180-degree acquisition produces reconstructed images of superior contrast and resolution, because the projection images of the heart from the opposite 180 degrees have poor spatial resolution and contrast due to greater distance and attenuation. Although studies have shown that the 180-degree acquisition can introduce artifacts, the 180-degree acquisition is more commonly used than the 360-degree for cardiac studies.

SPECT projection images are usually acquired in either a 64^2 or a 128^2 pixel format. Using too small a pixel format reduces the spatial resolution of the projection images and of the resultant reconstructed transverse images. When the 64^2 format is used, typically 60 or 64 projection images are acquired and, when a 128^2 format is chosen, 120 or 128 projection images are acquired. Using too few projections creates radial streak artifacts in the reconstructed transverse images.

The camera heads on older SPECT systems followed circular orbits around the patient while acquiring images. Circular orbits are satisfactory for SPECT imaging of the brain, but cause a loss of spatial resolution in body imaging because the circular orbit causes the camera head to be many centimeters away from the surface of the body during the anterior and posterior portions of its orbit (Fig. 22-3). Newer SPECT systems provide noncircular orbits (also called "body contouring") that keep the camera heads in close proximity to the surface of the body throughout the orbit. For some systems, the technologist specifies the noncircular orbit by placing the camera head as close as possible to the patient at several angles, from which the camera's computer determines the orbit. Other systems perform automatic body

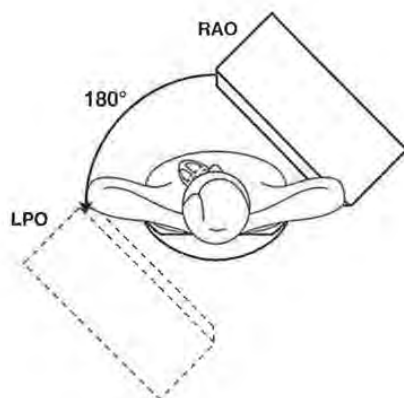


FIGURE 22-2. An 180-degree cardiac orbit.

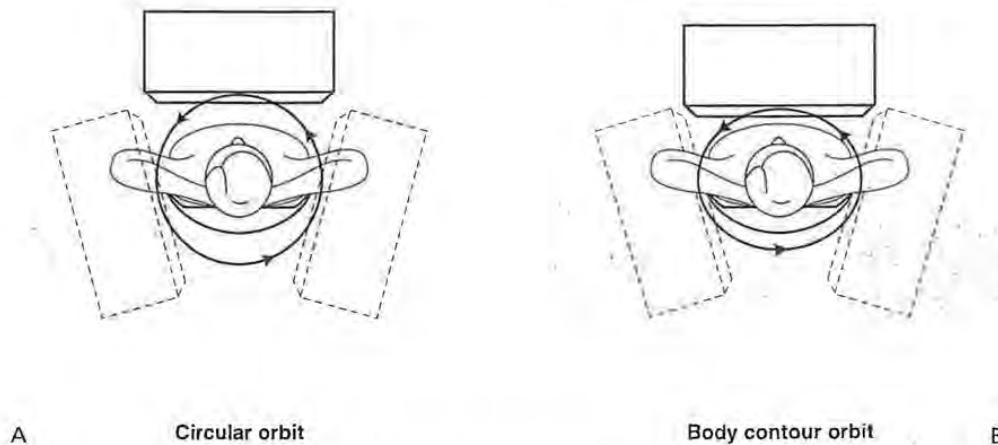


FIGURE 22-3. Circular (A) and body-contour (B) orbits.

contouring, using sensors on the camera heads to determine their proximity to the patient.

In brain SPECT, it is usually possible for the camera head to orbit with a much smaller radius than in body SPECT, thereby producing images of much higher spatial resolution. Unfortunately, in older cameras, the large distance from the physical edge of the camera head to the useful portion of the detector often made it impossible to orbit at a radius within the patient's shoulders while including the base of the brain in the images. These older systems were therefore forced to image the brain with an orbit outside the patient's shoulders, causing a significant loss of resolution. Most modern SPECT systems permit brain imaging with orbits within the patient's shoulders.

Transverse Image Reconstruction

After the projection images are acquired, they are usually corrected for axis-of-rotation misalignments and for nonuniformities. (These corrections are discussed below; see Quality Control in SPECT.) Following these corrections, transverse image reconstruction is performed using either filtered backprojection or iterative methods.

As described in Chapter 13, filtered backprojection consists of two steps. First, the projection images are mathematically filtered. Then, to form a particular transverse image, simple backprojection is performed of the row, corresponding to that transverse image, of each projection image. For example, the fifth row of each projection image is backprojected to form the fifth transverse image. A SPECT study produces transverse images covering the entire field of view (FOV) of the camera in the axial direction from each revolution of the camera head or heads.

Mathematical theory specifies that the ideal filter kernel, when displayed in frequency space, is the ramp filter (Fig. 22-4). However, the actual projection images contain considerable statistical noise. If they were filtered using a ramp filter kernel and then backprojected, the resultant transverse images would contain an unacceptable amount of statistical noise.

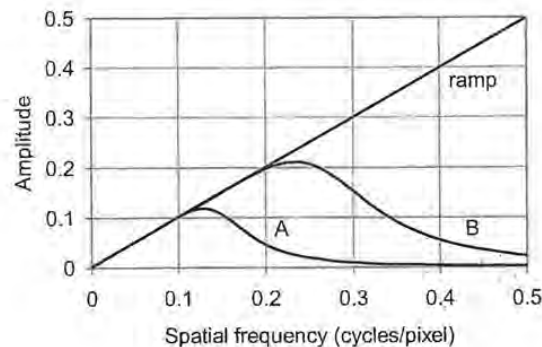


FIGURE 22-4. Typical filter kernels used for filtered backprojection. The kernels are shown in frequency space. Filter kernel A is a Butterworth filter of the 5th order with a critical frequency of 0.15 Nyquist, and filter kernel B is a Butterworth filter of the 5th order with a critical frequency of 0.27 Nyquist. Filter kernel A provides more smoothing than filter kernel B. A ramp filter, which provides no smoothing, is also shown.

In the spatial frequency domain, statistical noise predominates in the high-frequency portion. To smooth the projection images before backprojection, the ramp filter kernel is modified to “roll off” at higher spatial frequencies. Unfortunately, this reduces the spatial resolution of the projection images and thus of the reconstructed transverse images. A compromise must therefore be made between spatial resolution and statistical noise of the transverse images.

Typically, a different filter kernel is selected for each type of SPECT study; e.g., a different kernel would be used for HMPAO brain SPECT than would be used for Tl 201 myocardial perfusion SPECT. The choice of filter kernel for a particular type of study is determined by the amount of statistical noise in the projection images (mainly determined by the injected activity, collimator, and acquisition time per image) and their spatial resolution [determined by the collimator and the typical distances of the camera head(s) from the organ being imaged]. The preference of the interpreting physician regarding the appearance of the images also plays a role. Projection images of better spatial resolution and less quantum mottle require a filter with a higher spatial frequency cutoff to avoid unnecessary loss of spatial resolution in the reconstructed transverse images, whereas projection images of poorer spatial resolution and greater quantum mottle require a filter with a lower spatial frequency cutoff to avoid excessive quantum mottle in the reconstructed transverse images. Although the SPECT camera's manufacturer may suggest filters for specific imaging procedures, the filters are usually empirically optimized in each nuclear medicine laboratory. Figure 22-5 shows a SPECT image created using three different filter kernels, illustrating too much smoothing, proper smoothing, and no smoothing.

Filtered backprojection is computationally efficient. However, it is based on the assumption that the projection images are perfect projections of a three-dimensional object. As discussed in the previous chapter, this is far from true in nuclear medicine imaging, mainly because of attenuation of photons in the patient, inclusion of Compton scattered photons in the image, and the degradation of spatial resolution with distance from the collimator.

In SPECT, iterative reconstruction methods are increasingly being used instead of filtered backprojection. In iterative methods, an initial activity distribution in the patient is assumed. Then, projection images are calculated from the assumed activity distribution, using the known imaging characteristics of the scintillation camera.



FIGURE 22-5. SPECT images created by filtered backprojection. The projection images were filtered using the filter kernels shown in Fig. 22-4. **Left:** An image produced using filter kernel A, which exhibits a significant loss of spatial resolution. **Center:** An image produced using filter kernel B, which provides a proper amount of smoothing. **Right:** An image produced using the ramp filter, which shows good spatial resolution but excessive statistical noise.

The calculated projection images are compared with the actual projection images and, based on this comparison, the assumed activity distribution is adjusted. This process is repeated several times, with successive adjustments to the assumed activity distribution, until the calculated projection images approximate the actual projection images.

As was stated above, in each iteration projection images are calculated from the assumed activity distribution. The calculation of projection images uses the point spread function of the scintillation camera, which takes into account the decreasing spatial resolution with distance from the camera face. The point spread function can be modified to incorporate the effect of photon scattering in the patient. Furthermore, if a map of the attenuation characteristics of the patient is available, the calculation of the projection images can include the effects of attenuation. If this is done, iterative methods will partially compensate for the effects of decreasing spatial resolution with distance, photon scattering in the patient, and attenuation in the patient.

Iterative methods are computationally less efficient than filtered backprojection. However, the increasing speed of computers, the small image matrix sizes used in nuclear imaging, and development of computationally efficient algorithms have made iterative reconstruction feasible for SPECT.

Attenuation Correction in SPECT

X- or gamma rays that must traverse long paths through the patient produce fewer counts, due to attenuation, than do those from activity closer to the near surface of the patient. For this reason, transverse image slices of a phantom with a uniform activity distribution, such as a jug filled with a well-mixed solution of radionuclide, will show a gradual decrease in activity toward the center (Fig. 22-6 left). Attenuation effects are more severe in body SPECT than in brain SPECT.

Approximate methods are available for attenuation correction. One of the most common, the Chang method, assumes a constant attenuation coefficient throughout the patient. Approximate attenuation corrections can over- or undercompensate

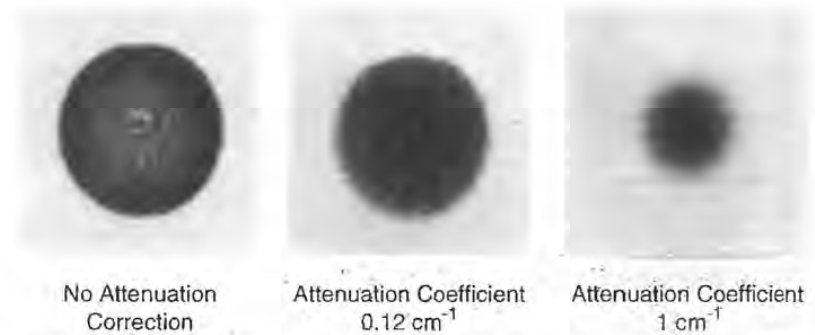


FIGURE 22-6. Attenuation correction. **Left:** A reconstructed transverse image slice of a cylindrical phantom containing a well-mixed radionuclide solution. This image shows a decrease in activity toward the center due to attenuation. (A small ring artifact, unrelated to the attenuation, is also visible in the center of the image.) **Center:** The same image corrected by the Chang method, using a linear attenuation coefficient of 0.12 cm^{-1} , demonstrating proper attenuation correction. **Right:** The same image, corrected by the Chang method using an excessively large attenuation coefficient.

for attenuation. If such a method is to be used, its proper functioning should be verified using phantoms before it is used in clinical studies.

Attenuation is not uniform throughout the patient and, in the thorax in particular, it is very nonuniform. These approximate methods cannot compensate for nonuniform attenuation.

Several manufacturers provide SPECT cameras with radioactive sources to measure the attenuation through the patient. The sources are used to acquire transmission data from projections around the patient. After acquisition, the transmission projection data are reconstructed to provide maps of tissue attenuation characteristics across transverse sections of the patient, similar to x-ray CT images. Finally, these attenuation maps are used during the SPECT image reconstruction process to provide attenuation-corrected SPECT images.

The transmission sources are available in several configurations. These include scanning collimated line sources that are used with parallel-hole collimators, arrays of fixed line sources used with parallel-hole collimators, and a fixed line source located at the focal point of a fan-beam collimator.

The transmission data are usually acquired simultaneously with the acquisition of the emission projection data, because performing the two separately poses significant problems in the spatial alignment of the two data sets. The radionuclide used for the transmission measurements is chosen to have primary gamma-ray emissions that differ significantly in energy from those of the radiopharmaceutical. Separate energy windows are used to differentiate the photons emitted by the transmission source from those emitted by the radiopharmaceutical. However, scattering of the higher energy photons in the patient and in the detector causes some crosstalk in the lower energy window.

In SPECT, attenuation correction using transmission sources has most often been applied in myocardial perfusion imaging, where attenuation artifacts can mimic perfusion defects. To date, however, clinical studies evaluating attenuation correction using transmission sources have not been uniformly favorable. At this

time, attenuation correction in SPECT using transmission sources is promising, but it is still under development and its ultimate clinical utility is not yet proven.

Generation of Coronal, Sagittal, and Oblique Images

The pixels from the transverse image slices may be reordered to produce coronal and sagittal images. For cardiac imaging, it is desirable to produce oblique images oriented either parallel (vertical and horizontal long axis images) or perpendicular (short axis images) to the long axis of the left ventricle. Because there is considerable anatomic variation among patients regarding the orientation of the long axis of the left ventricle, the long axis of the heart must be marked on the displayed image.

SPECT Collimators

The most commonly used collimator for SPECT is the high-resolution parallel-hole collimator. However, specialized collimators have been developed for SPECT. The fan-beam collimator, shown in Fig. 22-7, is a hybrid of the converging and the parallel-hole collimators. Because it is a parallel-hole collimator in the y-direction, each row of pixels in a projection image corresponds to a single transaxial slice of the subject. In the x-direction, it is a converging collimator, with spatial resolution-efficiency characteristics that are superior to those of a parallel-hole collimator (see Chapter 21, Fig. 21-13). Because a fan-beam collimator is a converging collimator in the cross-axial direction, its FOV decreases with distance from the collimator. For this reason, the fan-beam collimator is mainly used for brain SPECT; if the collimator is used for body SPECT, portions of the body are excluded from the FOV, creating artifacts in the reconstructed images.

Multihead SPECT Cameras

To reduce the limitations imposed on SPECT by collimation and limited time per view, camera manufacturers provide SPECT systems with two or three scintillation

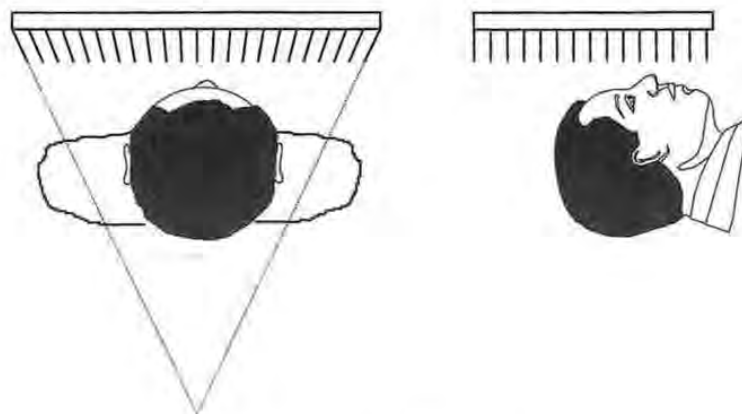


FIGURE 22-7. Fan-beam collimator.

camera heads that revolve about the patient (Fig. 22-1). The use of multiple camera heads permits the use of higher resolution collimators, for a given level of quantum mottle in the images, than would a single head system. However, the use of multiple camera heads imposes considerable technical difficulties on the manufacturer. It places severe requirements on the electrical and mechanical stability of the camera heads. Furthermore, the Y-offsets and X- and Y-magnification factors of all the heads must be precisely matched throughout rotation. Early multihead systems failed to achieve acceptance because of the difficulty of matching these factors. Today's multihead systems are more mechanically and electronically stable than earlier systems, providing high-quality tomographic images for a variety of clinical applications.

Multihead scintillation cameras are available in several configurations. Double-head cameras fixed in a 180-degree configuration are good for head and body SPECT and whole-body planar scans. Triple-head, fixed-angle cameras are good for head and body SPECT, but less suitable for whole-body planar scans because of the limited widths of the crystals. Double-head, variable-angle cameras are highly versatile, capable of head and body SPECT and whole-body planar scans in the 180-degree configuration and cardiac SPECT in the 90-degree configuration. (The useful portion of the NaI crystal does not extend all the way to the edge of a camera head. If the two camera heads are placed at an angle of exactly 90 degrees to each other, both heads cannot be close to the patient without parts of the patient being outside of the fields of view. For this reason, one manufacturer provides SPECT acquisitions with the heads at a 76-degree angle to each other.)

Performance

Spatial Resolution

The spatial resolution of a SPECT system can be measured by acquiring a SPECT study of a line source, such as a capillary tube filled with a solution of technetium (Tc) 99m, placed parallel to the axis of rotation. The National Electrical Manufacturers Association (NEMA) specifies a cylindrical plastic water-filled phantom, 22 cm in diameter, containing three line sources (Fig. 22-8, left) for measuring spatial resolution. The full-width-at-half-maximum (FWHM) measures of the line sources are determined from the reconstructed transverse images (Fig. 22-8, right). A ramp filter is used in the filtered backprojection, so that the filtering does not reduce the spatial resolution. The NEMA spatial resolution measurements are primarily determined by the collimator used. The tangential resolution for the peripheral sources (typically 7 to 8 mm FWHM for low-energy high-resolution and ultra-high-resolution parallel-hole collimators) is much superior to the central resolution (typically 9.5 to 12 mm). The tangential resolution for the peripheral sources is better than the radial resolution (typically 9.4 to 12 mm) for the peripheral sources.

These FWHMs measured using the NEMA protocol, while providing a useful index of ultimate system performance, are not necessarily representative of clinical performance, because these spatial resolution studies can be acquired using longer imaging times and closer orbits than would be possible in a patient. Patient studies may require the use of lower resolution (higher efficiency) collimators than the one used in the NEMA measurement to obtain adequate image statistics. In addition, the filters used before backprojection for clinical studies have lower spatial fre-

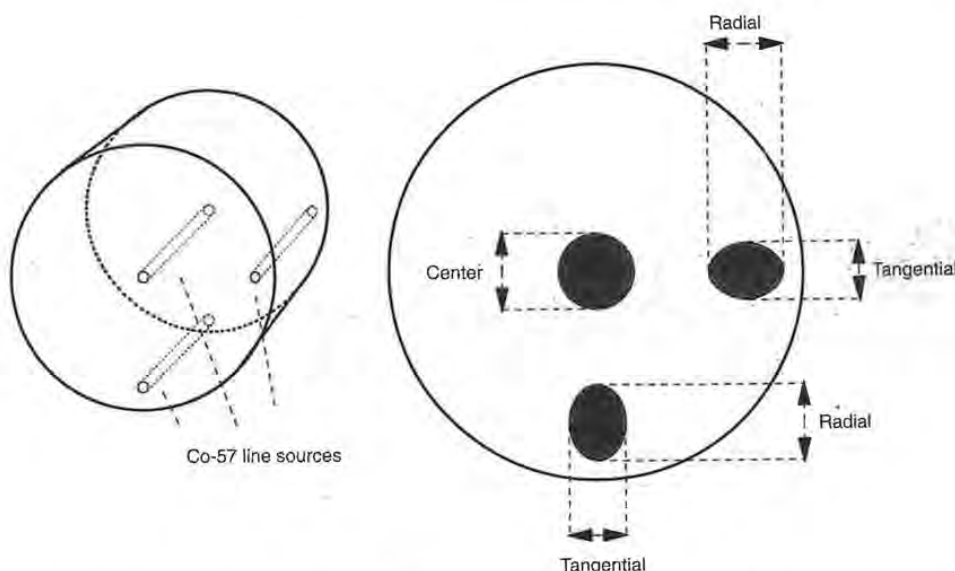


FIGURE 22-8. National Electrical Manufacturers Association (NEMA) phantom for evaluating the spatial resolution of a SPECT camera (**left**). The phantom is a 22 cm diameter plastic cylinder, filled with water and containing three Co 57 line sources. One line source lies along the central axis and the other two are parallel to the central line source, 7.5 cm away. A SPECT study is acquired with a camera radius of rotation [distance from collimator to axis of rotation (AOR)] of 15 cm. The spatial resolution is measured from reconstructed transverse images (**right**). Horizontal and vertical profiles are taken through the line sources and the full-width-at-half-maximum (FWHM) measures of these line spread functions (LSFs) are determined. The average central resolution (FWHM of the central LSF) and the average tangential and radial resolutions (determined from the FWHMs of the two peripheral sources as shown) are determined. (Adapted from Performance Measurements of Scintillation Cameras, National Electrical Manufacturers Association, 1986.)

quency cutoffs than do the ramp filters used in NEMA spatial resolution measurements. The NEMA spatial resolution measurements fail to show the advantage of SPECT systems with two or three camera heads; double- and triple-head cameras will permit the use of higher resolution collimators for clinical studies than will single-head cameras.

The spatial resolution deteriorates as the radius of the camera orbit increases. For this reason, brain SPECT produces images of much higher spatial resolution than body SPECT. It is essential that the SPECT camera heads orbit the patient as closely as possible. Also, noncircular orbits (see Design and Principles of Operation, above) provide better resolution than circular orbits.

Comparison of SPECT to Conventional Planar Scintillation Camera Imaging

In theory, SPECT should produce spatial resolution similar to that of planar scintillation camera imaging. In clinical imaging, its resolution is usually slightly worse. The camera head is usually closer to the patient in conventional planar imaging than in SPECT. Also, the short time per view of SPECT may mandate the use of a lower resolution collimator to obtain adequate numbers of counts.

In planar nuclear imaging, radioactivity in tissues in front of and behind an organ or tissue of interest causes a reduction in contrast. Furthermore, if the activity in these overlapping structures is not uniform, the pattern of this activity distribution is superimposed on the activity distribution in the organ or tissue of interest. As such, it is a source of structural noise that impedes the ability to discern the activity distribution in the organ or tissue of interest. The main advantage of SPECT over conventional planar nuclear imaging is the markedly improved contrast and reduced structural noise produced by eliminating the activity in overlapping structures. SPECT also offers the promise of partial correction for the effects of attenuation and of the scattering of photons in the patient.

Quality Control in SPECT

Although a technical quality control program is important in planar nuclear imaging, it is critical in SPECT. Equipment malfunctions or maladjustments that would not noticeably affect planar images can markedly degrade the spatial resolution of SPECT images and produce significant artifacts, some of which may mimic pathology. Upon installation, a SPECT camera should be tested by a medical physicist. Following acceptance testing, a quality control program should be established to ensure that the system's performance remains comparable to its initial performance.

X- and Y-Magnification Factors and Multienergy Spatial Registration

The X- and Y-magnification factors, often called X and Y gains, relate distances in the object being imaged, in the x and y directions, to the numbers of pixels between the corresponding points in the resultant image. The X-magnification factor is determined from a digital image of two point sources placed against the camera's collimator a known distance apart along a line parallel to the X-axis:

$$X_{\text{mag}} = \frac{\text{Actual distance between centers of point sources}}{\text{Number of pixels between centers of point sources}}$$

The Y-magnification factor is determined similarly but with the sources parallel to the y-axis. The X- and Y-magnification factors should be equal. If they are not, the projection images will be distorted in shape, as will be coronal, sagittal, and oblique images. (The transverse images, however, will not be distorted.)

The multienergy spatial registration, described in Chapter 21, is a measure of the camera's ability to maintain the same image magnification, regardless of the energies of the x- or gamma rays forming the image. It is important in SPECT when imaging radionuclides such as Ga 67 and In 111, which emit useful photons of more than one energy. It is also important because uniformity and axis of rotation corrections, to be discussed shortly, that are determined with one radionuclide will only be valid for others if the multienergy spatial registration is correct.

Alignment of Projection Images to Axis-of-Rotation (COR Calibration)

The *axis of rotation* (AOR) is an imaginary reference line about which the head or heads of a SPECT camera rotate. If a radioactive line source were placed on the AOR, each projection image would depict a vertical straight line near the center of



FIGURE 22-9. Center-of-rotation (COR) misalignment. Small misalignments cause blurring (**center**), whereas large misalignments cause point sources to appear as “tiny doughnut” artifacts (**right**).

the image; this projection of the AOR into the image is called the *center of rotation* (COR). Ideally, the COR is aligned with the center, in the x-direction, of each projection image. However, there may be a misalignment of the COR with the center of the projection images. This misalignment may be mechanical; for example, the camera head may not be exactly centered in the gantry. It can also be electronic. The misalignment may be the same amount in all projection images from a single camera head, or it may vary with the angle of the projection image.

If a COR misalignment is not corrected, it causes a loss of spatial resolution in the resultant transverse images. If the misalignment is large, it can cause a point source to appear as a tiny “doughnut” (Fig. 22-9). (These doughnut artifacts are not seen in clinical images; they are visible only reconstructed images of point or line sources. The doughnut artifacts caused by COR misalignment are not centered in the image and so can be distinguished from the ring artifacts caused by nonuniformities.) The COR alignment is assessed by placing a point source or line source in the camera field of view, acquiring a set of projection images, and analyzing these images using the SPECT system’s computer. If a line source is used, it is placed parallel to the AOR. The COR misalignment may be corrected by shifting each image in the x-direction by the proper number of pixels prior to filtered backprojection. When a line source is used, the COR correction can be performed separately for each slice. If the COR misalignment varies with camera head angle, instead of being constant for all projection images, it can only be corrected if the computer permits angle-by-angle corrections. Separate assessments of the COR correction must be made for different collimators and, on some systems, for different camera zoom factors and image formats (e.g., 64^2 versus 128^2). The COR correction determined using one radionuclide will only be valid for other radionuclides if the multienergy spatial registration is correct.

Uniformity

The uniformity of the camera head or heads is important; nonuniformities that are not apparent in low-count daily uniformity studies can cause significant artifacts in SPECT. The artifact caused by a nonuniformity appears in transverse images as a ring centered about the AOR (Fig. 22-10).

Multihead SPECT systems can produce partial ring artifacts when projection images are not acquired by all heads over a 360-degree arc. Clinically, ring artifacts are most apparent in high count-density studies, such as liver images. However, ring artifacts may be most harmful in studies such as Tl-201 myocardial perfusion images in which, due to poor counting statistics and large variations in count density, they may not be recognized and thus may lead to misinterpretation.



FIGURE 22-10. Image of a cylinder filled with a uniform radionuclide solution, showing a ring artifact due to a nonuniformity. Artifact is the dark ring toward the center.

As mentioned in the previous chapter, the primary intrinsic causes of nonuniformity are (a) spatial nonlinearities, which “stretch” the image in some areas, reducing the local count density, and “compress” other areas of the image, increasing the count density; and (b) local variations in the efficiency of light collection, which produce systematic position-dependent variations in the amplitudes of the energy signals and thus local variations in the fraction of counts rejected by the energy discrimination circuits. All modern SPECT cameras incorporate digital circuits that apply correction factors, obtained from internal “lookup” tables, to the X, Y, and energy signals from each interaction. However, these correction circuits cannot correct nonuniformity due to local variations in detection efficiency, such as dents or manufacturing defects in the collimators. If not too severe, nonuniformities of this type can be largely corrected. A very high count uniformity image must be acquired. The ratio of the average pixel count to the count in a specific pixel in this image serves as a correction factor for that pixel. Following the acquisition of a projection image during a SPECT study, each pixel of the projection image is multiplied by the appropriate correction factor before COR correction and filtered backprojection. For the high-count uniformity image, at least 30 million counts should be collected for 64^2 images and 120 million counts for a 128^2 format. These high-count uniformity images should be collected every 1 or 2 weeks. Separate correction images must be acquired for each camera head. For cameras from some manufacturers, separate correction images must be acquired for each radionuclide. The effectiveness of a camera’s correction circuitry and of the use of high-count flood correction images can be tested by acquiring a SPECT study of a large plastic jug filled with a well-mixed solution of Tc 99m and examining the transverse images for ring artifacts.

Camera Head Tilt

The camera head or heads must be exactly parallel to the AOR. If they are not, a loss of spatial resolution and contrast will result from out-of-slice activity being backprojected into each transverse image slice (Fig. 22-11). The loss of resolution and contrast in each transverse image slice will be less toward the center of the slice and greatest toward the edges of the image. If the axis of rotation of the camera is aligned to be level when the camera is installed and there is a flat surface on the

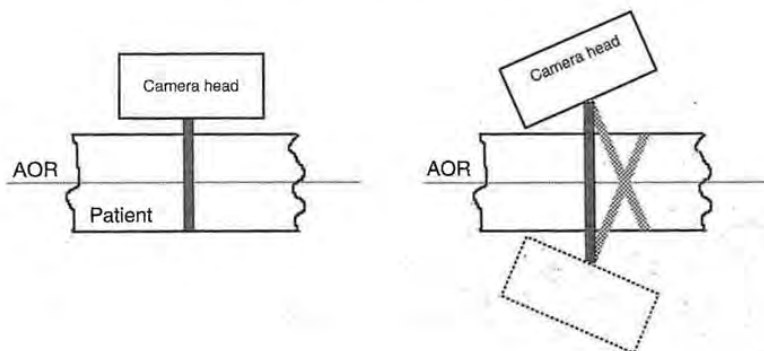


FIGURE 22-11. Head tilt. The camera head on the left is parallel to the axis of rotation (AOR), causing the counts collected in a pixel of the projection image to be backprojected into the same slice of the patient. The camera head (**right**) is tilted, causing counts from activity outside of a transverse slice (*lighter shading*) to be backprojected into the transverse slice (*darker shading*).

camera head that is parallel to the collimator face, a bubble level may be used to test for head tilt. Some SPECT cameras require the head tilt to be manually adjusted for each acquisition, whereas other systems set it automatically. The accuracy of the automatic systems should be periodically tested. A more reliable method than the bubble level is to place a point source in the camera FOV, centered in the axial (y) direction, but near the edge of the field in the transverse (x) direction. A series of projection images is then acquired. If there is head tilt, the position of the point source will vary in the y-direction from image to image. Proper head tilt is easily evaluated from cine viewing of the projection images.

SPECT Phantoms

There are commercially available phantoms (Fig. 22-12) that may be filled with a solution of Tc 99m or other radionuclide and used to evaluate system performance.

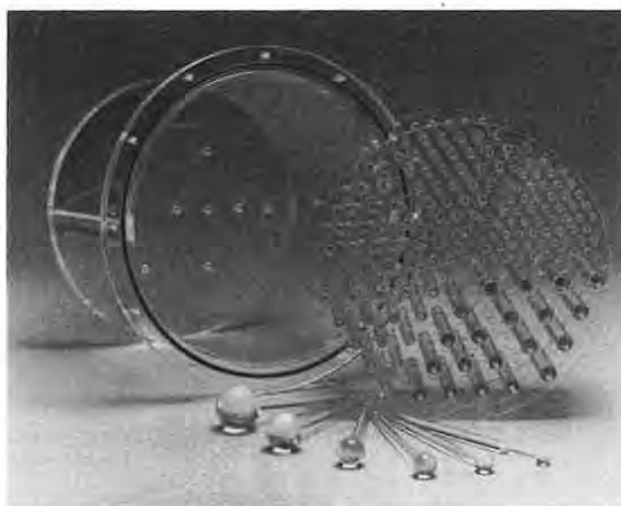


FIGURE 22-12. "Jaszczak" phantom for testing SPECT systems. (Courtesy of Data Spectrum Corp.)

TABLE 22-1. RECOMMENDED SCHEDULE FOR ROUTINE QUALITY CONTROL TESTING OF A SCINTILLATION CAMERA

Test	Frequency	Notes
Set and check energy window(s)	Daily and before imaging different radionuclide	Point source of radionuclide to be imaged
Cine review of projection images and/or review of sinogram (S) ^a	After each clinical SPECT study	Check for patient motion
Visual inspection of collimators for damage	Daily and when changing collimators	If new damage found, acquire a high-count uniformity image
Extrinsic or intrinsic low-count uniformity images of all camera heads	Daily	2 to 6 million counts, depending on effective area of camera head
High count-density extrinsic or intrinsic uniformity images of all camera heads (S)	1 or 2 weeks	30 million counts for 64 ² images and 120 million counts for 128 ²
Spatial resolution check with bar or hole pattern	1 or 2 weeks	
Center of rotation (S)	1 or 2 weeks	Point or line source, as recommended by manufacturer
Efficiency of each camera head	Quarterly	
Reconstructed cylindrical phantom uniformity (S)	Monthly	Cylindrical phantom filled with Tc-99m solution
Point source reconstructed spatial resolution (S)	Quarterly	Point source
Reconstructed SPECT phantom (S)	Quarterly	Using a phantom such as the one shown in Fig. 22-10
Pixel size check	Quarterly	Two point sources
Head tilt angle check (S)	Quarterly	Bubble level or point source
Extrinsic uniformity images of all collimators not tested above	Quarterly or semiannually	Planar source; high count density images of all collimators used for SPECT
Multienergy spatial registration	Quarterly, semiannually, or annually	Ga-67 point sources
Count rate performance	Annually	Tc-99m source

Note: The results of these tests are to be compared with baseline values, typically determined during acceptance testing. If the manufacturer recommends additional tests or more frequent testing, the manufacturer's recommendations should take precedence. For tests whose frequencies are listed as "1 or 2 weeks" or "quarterly or semiannually," it is recommended that the tests be performed initially at the higher frequency, but the frequency be reduced if the measured parameters prove stable. Tests labeled (S) need not be performed for cameras used only for planar imaging.

^aA sinogram is an image containing projection data corresponding to a single transaxial image of the patient. Each row of pixels in the sinogram is the row, corresponding to that transaxial image, of one projection image.

SPECT, single photon emission computed tomography.

These phantoms are very useful for the semiquantitative assessment of spatial resolution, image contrast, and uniformity. They are used for acceptance testing of new systems and periodic testing, typically quarterly, thereafter. Table 22-1 provides a suggested schedule for a SPECT quality-control program.

22.2 POSITRON EMISSION TOMOGRAPHY (PET)

Positron emission tomography (PET) generates images depicting the distributions of positron-emitting nuclides in patients. Figure 22-13 shows a PET scanner. In the typical scanner, several rings of detectors surround the patient. PET scanners use *annihilation coincidence detection* (ACD) instead of collimation to obtain projections of the activity distribution in the subject. The PET system's computer then reconstructs the transverse images from the projection data, as does the computer of an x-ray CT or SPECT system. Modern PET scanners are multislice devices, permitting the simultaneous acquisition of as many as 45 slices over 16 cm of axial distance. Although there are many positron-emitting radiopharmaceuticals, the clinical importance of PET today is largely due to its ability to image the radiopharmaceutical fluorine-18 fluorodeoxyglucose (FDG), a glucose analog used for differentiating malignant neoplasms from benign lesions, staging malignant neoplasms, differentiating severely hypoperfused but viable myocardium from scar, and other applications.



FIGURE 22-13. A commercial PET scanner. (Courtesy of Siemens Medical Systems, Nuclear Medicine Group.)

Design and Principles Of Operation

Annihilation Coincidence Detection

Positron emission is a mode of radioactive transformation and was discussed in Chapter 18. Positrons emitted in matter lose most of their kinetic energy by causing ionization and excitation. When a positron has lost most of its kinetic energy, it interacts with an electron by *annihilation* (Fig. 22-14, left). The entire mass of the electron-positron pair is converted into two 511-keV photons, which are emitted in nearly opposite directions. In solids and liquids, positrons travel only very short distances before annihilation.

If both of these annihilation photons interact with detectors, the annihilation occurred close to the line connecting the two interactions (Fig. 22-14, right). Circuitry within the scanner identifies interactions occurring at nearly the same time, a process called annihilation coincidence detection (ACD). The circuitry of the scanner then determines the line in space connecting the locations of the two interactions. Thus, ACD establishes the trajectories of detected photons, a function performed by collimation in SPECT systems. However, the ACD method is much less wasteful of photons than collimation. Additionally, ACD avoids the degradation of spatial resolution with distance from the detector that occurs when collimation is used to form projection images.

True, Random, and Scatter Coincidences

A *true coincidence* is the simultaneous interaction of emissions resulting from a single nuclear transformation. A *random coincidence* (also called an *accidental* or *chance coincidence*), which mimics a true coincidence, occurs when emissions from different nuclear transformations interact simultaneously with the detectors (Fig. 22-15). A *scatter coincidence* occurs when one or both of the photons from a single annihilation are scattered, but both are detected (Fig. 22-15). A scatter coincidence is a

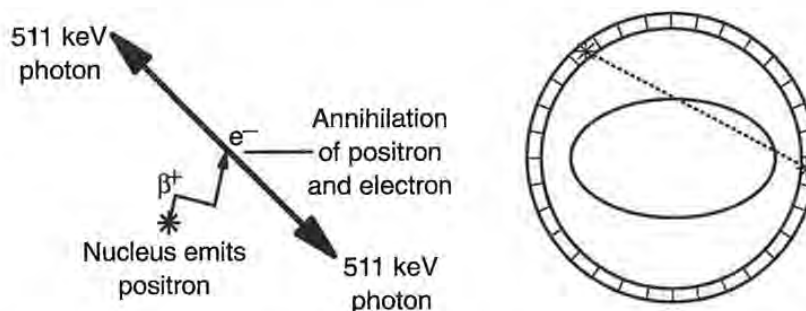


FIGURE 22-14. Annihilation coincidence detection (ACD). When a positron is emitted by a nuclear transformation, it scatters through matter losing energy. After it loses most of its energy, it annihilates with an electron, resulting in two 511-keV photons that are emitted in nearly opposite directions (**left**). When two interactions are simultaneously detected within a ring of detectors surrounding the patient (**right**), it is assumed that an annihilation occurred on the line connecting the interactions. Thus, ACD, by determining the path of the detected photons, performs the same function for the PET scanner as does the collimator of a scintillation camera.

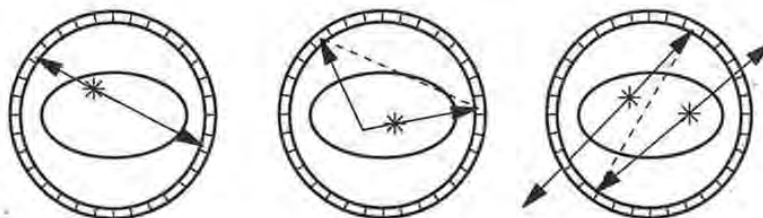


FIGURE 22-15. True coincidence (**left**), scatter coincidence (**center**), and random (accidental) coincidence (**right**). A scatter coincidence is a true coincidence, because it is caused by a single nuclear transformation, but results in a count attributed to the wrong line-of-response (*dashed line*). The random coincidence is also attributed to the wrong line of response.

true coincidence, because both interactions result from a single positron annihilation. Random coincidences and scatter coincidences result in misplaced coincidences, because they are assigned to lines of response that do not intersect the actual locations of the annihilations. They are therefore sources of noise, whose main effect is to reduce image contrast.

Design of a PET Scanner

Scintillation crystals coupled to photomultiplier tubes (PMTs) are used as detectors in PET; the low intrinsic efficiencies of gas-filled and semiconductor detectors for detecting 511-keV photons make them impractical for use in PET. The signals from the PMTs are processed using pulse mode (the signals from each interaction are processed separately from those of other interactions) to create signals identifying the position, deposited energy, and time of each interaction. The energy signal is used for energy discrimination to reduce mispositioned events due to scatter and the time signal is used for coincidence detection.

In early PET scanners, each scintillation crystal was coupled to a single PMT. In this design, the size of the individual crystal largely determined the spatial resolution of the system; reducing the size (and therefore increasing the number of crystals) improved the resolution. It became increasingly costly and impractical to pack more and more smaller PMTs into each detector ring. Modern designs couple larger crystals to more than one PMT; one such design is shown in Fig. 22-16. The relative magnitudes of the signals from the PMTs coupled to a single crystal are used to determine the position of the interaction in the crystal, as in a scintillation camera.

The scintillation material must emit light very promptly to permit true coincident interactions to be distinguished from random coincidences and to minimize dead-time count losses at high interaction rates. Also, to maximize the counting efficiency, the material must have a high linear attenuation coefficient for 511-keV photons. Most PET systems today use crystals of bismuth germanate ($\text{Bi}_4\text{Ge}_3\text{O}_{12}$, abbreviated BGO). The light output of BGO is only 12% to 14% of that of $\text{NaI}(\text{Tl})$, but its greater density and average atomic number give it a much higher efficiency in detecting 511-keV annihilation photons. Light is emitted rather slowly from BGO (decay constant of 300 nsec), which contributes to dead-time count

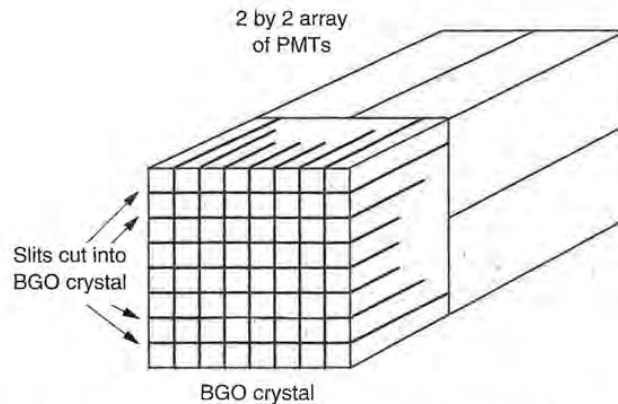


FIGURE 22-16. A technique for coupling scintillation crystals to photomultiplier tubes (PMTs). The relative heights of the pulses from the four PMTs are used to determine the position of each interaction in the crystal. The thick (3 cm) crystal is necessary to provide reasonable detection efficiency for the 511-keV annihilation photons. The slits cut in the bismuth germanate crystal form light pipes, limiting the spread of light from interactions in the front portion of the crystal, which otherwise would reduce the spatial resolution. This design permits four PMTs to serve 64 detector elements.

losses and random coincidences at high interaction rates. Several new inorganic scintillators are being investigated as possible replacements for BGO. Two of the most promising of these are lutetium oxyorthosilicate ($\text{Lu}_2\text{SiO}_4\text{O}$, abbreviated LSO) and gadolinium oxyorthosilicate ($\text{Gd}_2\text{SiO}_4\text{O}$, abbreviated GSO), both activated with cerium. Their attenuation properties are nearly as favorable as those of BGO and their much faster light emission produces better performance at high interaction rates, especially in reducing dead-time effects and in discriminating between true and random coincidences. Their higher conversion efficiencies may produce improved intrinsic spatial resolution and rejection of scatter. A disadvantage of LSO is that it is slightly radioactive; about 2.6% of naturally occurring lutetium is radioactive $\text{Lu } 176$, which has a half-life of 38 billion years, producing about 295 nuclear transformations per second in each cubic centimeter of LSO. $\text{NaI}(\text{Tl})$ is used as the scintillator in some less expensive PET systems. The properties of BGO, LSO(Ce), and GSO(Ce) are contrasted with those of $\text{NaI}(\text{Tl})$ in Table 22-2.

The energy signals from the detectors are sent to energy discrimination circuits, which can reject events in which the deposited energy differs significantly from 511 keV to reduce the effect of photon scattering in the patient. However, some annihilation photons that have escaped the patient without scattering interact with the detectors by Compton scattering, depositing less than 511 keV. An energy discrimination window that encompasses only the photopeak rejects these valid interactions as well as photons that have scattered in the patient. The energy window can be set to encompass only the photopeak, with maximal rejection of scatter, but also reducing the number of valid interactions detected; or the window can include part

TABLE 22-2. PROPERTIES OF SEVERAL INORGANIC SCINTILLATORS OF INTEREST IN PET

Scintillator	Decay constant (nsec)	Peak wavelength (nm)	Atomic numbers	Density (g/cm ³)	Attenuation coefficient 511 keV (cm ⁻¹)	Conversion efficiency relative to NaI
NaI(Tl)	230	415	11,53	3.67	0.343	100%
BGO	300	460	83,32,8	7.17	0.964	12–14%
GSO(Ce)	56	430	64,14,8	6.71	0.704	41%
LSO(Ce)	40	420	71,14,8	7.4	0.870	75%

Decay constants, peak wavelengths of emitted light, densities, and conversion efficiencies of BGO, GSO, and LSO from Ficke DC, Hood JT, Ter-Pogossian MM. A spheroid positron emission tomograph for brain imaging: a feasibility study. *J Nucl Med* 1996;37:1222.

BGO, bismuth germanate; GSO, gadolinium oxyorthosilicate; LSO, lutetium oxyorthosilicate; PET, positron emission tomography.

of the Compton continuum, increasing the sensitivity, but also increasing the number of scattered photons detected.

The time signals of interactions not rejected by the energy discrimination circuits are used for coincidence detection. When a coincidence is detected, the circuitry or computer of the scanner determines a line in space connecting the two interactions (Fig. 22-14). This is called a *line of response*. The number of coincidences detected along each line of response is stored in the memory of the computer. Figure 22-17 compares the acquisition of projection data by SPECT and PET systems. Once data acquisition is complete, the computer uses the projection data to produce transverse images of the radionuclide distribution in the patient, as in x-ray CT or SPECT.

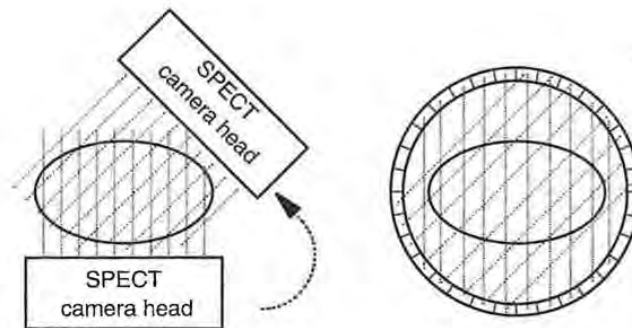


FIGURE 22-17. Acquisition of projection data by a SPECT system (left) and a PET system (right), showing how each system collects data for two projections. However, the PET system collects data for all projections simultaneously, whereas the scintillation camera head of the SPECT system must move to collect data from each projection angle.

Timing of Interactions and Detection of Coincidences

To detect coincidences, the times of individual interactions must be compared. However, interactions themselves cannot be directly detected; instead, the signals caused by interactions are detected. In scintillators, the emission of light after an interaction is characterized by a relatively rapid increase in intensity followed by a more gradual decrease. A timing circuit connected to each detector must designate one moment that follows each interaction by a constant time interval. The time signal is determined from the leading edge of the electrical signal from the PMTs connected to a detector crystal because the leading edge is steeper than the trailing edge. When the time signals from two detectors occur within a selected time interval called the *time window*, a coincidence is recorded. A typical time window for a system with BGO detectors is 12 nsec.

True Versus Random Coincidences

The rate of random coincidences between any pair of detectors is

$$R_{\text{random}} = 2\tau S_1 S_2 \quad [22-1]$$

where τ is the coincidence time window and S_1 and S_2 are the actual count rates of the detectors, often called *singles rates*. The time window is the time interval following an interaction in one detector in which an interaction in the other detector is considered to be a coincidence. The ratio of random to true coincidences increases with increasing activity in the subject and decreases as the time window is shortened. Unfortunately, there is a limit to how small the time window can be. If the time window is made too short, some true coincidences are also rejected because of imprecision in the timing of interactions. Scintillation materials that emit light promptly permit the use of shorter time windows and thus better discrimination between true and random coincidences.

Although individual random coincidences cannot be distinguished from true coincidences, methods of correction are available. The number of random coincidences along each line of response can be measured by adding a time delay to one of the two timing signals used for coincidence detection so that no true coincidences are detected. Alternatively, the number of random coincidences may be calculated from the single-event count rates of the detectors, using Equation 22-1. The number of random coincidences is then subtracted from the number of true plus random coincidences.

Scatter Coincidences

As previously mentioned, a scatter coincidence occurs when one or both of the photons from an annihilation are scattered in the patient and both are detected (Fig. 22-15). The fraction of scatter coincidences is dependent on the amount of scattering material and thus is less in head than body imaging. Because scatter coincidences are true coincidences, reducing the activity administered to the patient, reducing the time window, or using a scintillator with faster light emission does not reduce the scatter coincidence fraction. Furthermore, the corrections for random coincidences do not compensate for scatter coincidences. The energy discrimination circuits of the PET scanner reject some scatter coincidences. Unfortunately, the

relatively low conversion efficiency (Table 22-2) of BGO produces poor energy resolution, limiting the ability of energy discrimination to reject scatter. The fraction of scatter coincidences can also be reduced by two-dimensional data acquisition and axial collimation, as described in the following section. Approximate methods are available for scatter correction.

Two- and Three-Dimensional Data Acquisition

In two-dimensional (slice) data acquisition, coincidences are detected and recorded within each detector ring or small groups of adjacent detector rings. PET scanners designed for two-dimensional acquisition have thin annular collimators, typically made of tungsten, to prevent most radiation emitted by activity outside a transaxial slice from reaching the detector ring for that slice (Fig. 22-18). The fraction of scatter coincidences is greatly reduced in PET systems using two-dimensional data acquisition and axial collimation because of the geometry. Consider an annihilation occurring within a particular detector ring with the initial trajectories of the annihilation photons toward the detectors. If either of the photons scatters, it is likely that the new trajectory of the photon will cause it to miss the detector ring, thereby preventing a scatter coincidence. Furthermore, most photons from out-of-slice activity are absorbed by the axial collimators.

In two-dimensional data acquisition, coincidences within one or more pairs of adjacent detector rings may be added to improve the sensitivity (Fig. 22-19). The data from each pair of detector rings are added to that of the slice midway between the two rings. For example, if N is the number of a particular detector ring, coincidences between rings $N - 1$ and $N + 1$ are added to the projection data for ring N . For further sensitivity, coincidences between rings $N - 2$ and $N + 2$ can also be added to those of ring N . After the data from these pairs of detector rings are added, a standard two-dimensional reconstruction is performed to create a transverse image for this slice. Similarly, coincidences between two immediately adjacent detector rings, i.e., between rings N and $N + 1$, can be used to generate a transverse image

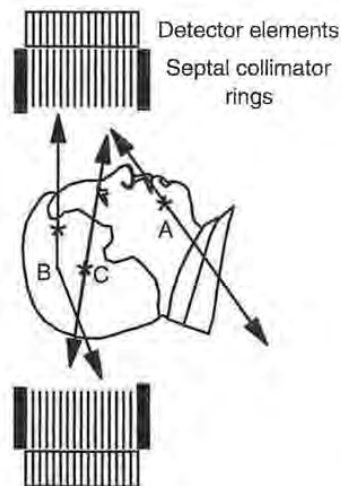


FIGURE 22-18. Side view of PET scanner illustrating two-dimensional data acquisition. The collimator rings prevent photons from activity outside the field of view (A) and most scattered photons (B) from causing counts in the detectors. However, many valid photon pairs (C) are also absorbed.

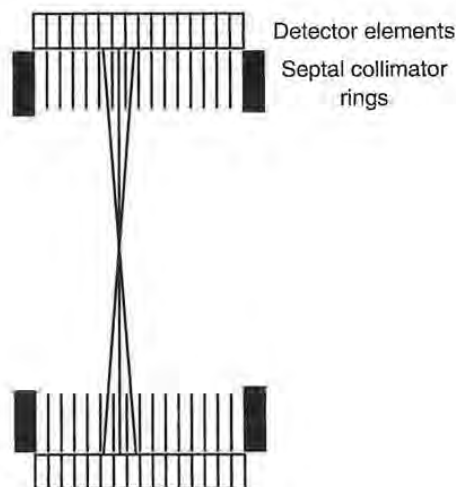


FIGURE 22-19. Side view of PET scanner performing two-dimensional data acquisition, showing lines of response for a single slice. Cross-ring coincidences have been added to those occurring within the ring of detector elements. This increases the number of coincidences detected, but causes a slight loss of axial spatial resolution that increases with distance from the axis of the scanner.

halfway between these rings. For greater sensitivity, coincidences between rings $N-1$ and $N+2$ can be added to these data. However, increasing the number of adjacent rings used in two-dimensional acquisition reduces the axial spatial resolution.

In three-dimensional (volume) data acquisition, axial collimators are not used and coincidences are detected between many or all detector rings (Fig. 22-20). Three-dimensional acquisition greatly increases the number of true coincidences detected and may permit smaller activities to be administered to patients. There are disadvantages to three-dimensional data acquisition. For the same administered activity, the greatly increased interaction rate increases the random coincidence fraction and the dead-time count losses. Thus, three-dimensional acquisition may

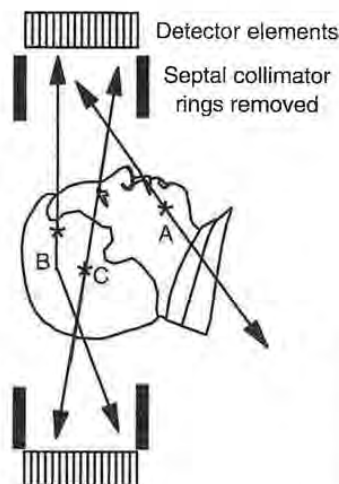


FIGURE 22-20. Side view of PET scanner illustrating three-dimensional data acquisition. Without axial collimator rings, interactions from activity outside the field of view (A) and scattered photons (B) are greatly increased, increasing the dead time, random coincidence fraction, and scatter coincidence fraction. However, the number of valid photon pairs (C) detected is also greatly increased.

require less activity to be administered, compared to two-dimensional image acquisition. Furthermore, the scatter coincidence fraction is much larger and the number of interactions from activity outside the FOV is greatly increased. (Activity out of the FOV causes few true coincidences, but increases the rate of random coincidences detected and dead-time count losses.) Thus, three-dimensional acquisition is likely to be most useful in low-scatter studies, such as pediatric and brain studies, and low-activity studies. Some PET systems are equipped with retractable axial collimators, permitting them to perform two- or three-dimensional acquisition.

Transverse Image Reconstruction

After the projection data are acquired, the data for each line of response is corrected for random coincidences, scatter coincidences, dead-time count losses, and attenuation (described below). Following these corrections, transverse image reconstruction is performed. For two-dimensional data acquisition, image reconstruction methods are similar to those used in SPECT. As in SPECT, either filtered backprojection or iterative reconstruction methods can be used. For three-dimensional data acquisition, special three-dimensional analytic or iterative reconstruction methods are required. These are beyond the scope of this chapter, but are discussed in the suggested readings listed at the end of the chapter.

An advantage of PET over SPECT is that, in PET, the correction for nonuniform attenuation can be applied to the projection data before reconstruction. In SPECT, the correction for nonuniform attenuation is intertwined with and complicates the reconstruction process.

Performance

Spatial Resolution

Modern whole-body PET systems achieve a spatial resolution slightly better than 5 mm FWHM of the line spread function (LSF) in the center of the detector ring when measured by the NEMA standard, *Performance Measurements of Positron Emission Tomographs*. There are three factors that primarily limit the spatial resolution of PET scanners: (a) the intrinsic spatial resolution of the detectors, (b) the distance traveled by the positrons before annihilation, and (c) the fact that the annihilation photons are not emitted in exactly opposite directions from each other. The intrinsic resolution of the detectors is the major factor determining the resolution in current scanners. In older systems in which the detectors consisted of separate crystals each attached to a single PMT, the size of the crystals determined the resolution.

The spatial resolution of a PET system is best in the center of the detector ring and decreases slightly (FWHM increases) with distance from the center. This occurs because of the considerable thickness of the detectors and because current PET systems cannot determine the depth in the crystal where an interaction occurs. Uncertainty in the depth of interaction causes uncertainty in the line of response for annihilation photons that strike the detectors obliquely. Photons emitted from the center of the detector ring can only strike the detectors head-on, but many of the photons emitted from activity away from the center strike the detectors from oblique angles (Fig. 22-21).

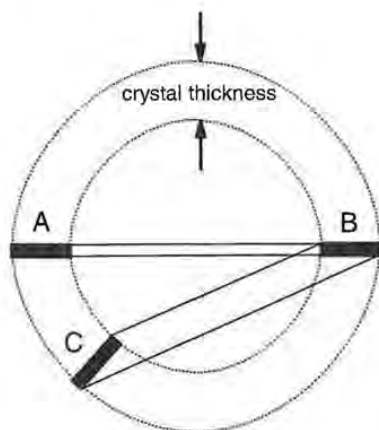


FIGURE 22-21. The cause of reduced spatial resolution with distance from the center of a PET scanner. Only three crystal segments in the detector ring are shown. For coincident interactions in detectors A and B, there is little uncertainty in the line-of-response. When coincidences occur in detectors A and C, there is greater uncertainty in the line of response. Using thinner crystals would reduce this effect, but would also reduce their intrinsic detection efficiency and thus the fraction of coincidences detected.

The distance traveled by the positron before annihilation also degrades the spatial resolution. This distance is determined by the maximal positron energy of the radionuclide and the density of the tissue. A radionuclide that emits lower energy positrons yields superior resolution. Table 22-3 lists the maximal energies of positrons emitted by radionuclides commonly used in PET. Activity in denser tissue yields higher resolution than activity in less dense tissue.

Although positrons lose nearly all of their momentum before annihilation, the positron and electron possess some residual momentum when they annihilate. Conservation of momentum predicts that the resultant photons will not be emitted in exactly opposite directions. This causes a small loss of resolution, which increases with the diameter of the detector ring.

The spatial resolution of a PET scanner, in clinical use, is likely to be somewhat worse than that measured using the NEMA protocol. The NEMA protocol specifies filtered backprojection reconstruction using a ramp filter, which provides no smoothing. Clinical projection data contains considerable statistical noise, requiring the use of a filter providing smoothing, which in turn degrades spatial resolution.

TABLE 22-3. PROPERTIES OF POSITRON-EMITTING RADIONUCLIDES COMMONLY USED IN PET

Nuclide	Half-life (minutes)	Maximal positron energy (keV)	Notes
C-11	20.5	960	
N-13	10.0	1,198	
O-15	2.0	1,732	
F-18	110	634	Most common use is ^{18}F -fluorodeoxyglucose
Rb-82	1.2	3,356	Produced by Sr-82/Rb-82 generator

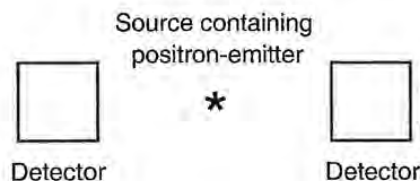


FIGURE 22-22. Point source of positron-emitting radionuclide, in air, midway between two identical detectors.

Efficiency in Annihilation Coincidence Detection

Consider a point source of a positron-emitting radionuclide in air midway between two identical detectors (Fig. 22-22) and assume that all positrons annihilate within the source. The true coincidence rate of the pair of detectors is

$$R_T = 2AG\varepsilon^2 \quad [22-2]$$

where A is the activity of the source, G is the geometric efficiency of either detector, and ε is the intrinsic efficiency of either detector. (Geometric and intrinsic efficiency are defined in Chapter 20.) Because the rate of true coincidences detected is proportional to the square of the intrinsic efficiency, maximizing the intrinsic efficiency is very important in PET. For example, if the intrinsic efficiency of a single detector for a 511-keV photon is 0.9, 81% of the annihilation photon pairs emitted toward the detectors have coincident interactions. However, if the intrinsic efficiency of a single detector is 0.1, only 1% of pairs emitted toward the detectors has coincident interactions. As mentioned previously, most PET systems today use crystals of BGO, because of its high attenuation coefficient. These crystals are typically about 3 cm thick. As mentioned above (see Two and Three Dimensional Data Acquisition), increasing the axial acceptance angle of annihilation photons greatly increases the efficiency.

Attenuation and Attenuation Correction in PET

Attenuation in PET differs from attenuation in SPECT, because both annihilation photons must escape the patient to cause a coincident event to be registered. The probability of both photons escaping the patient without interaction is the product of the probabilities of each escaping:

$$(e^{-\mu x}) \cdot (e^{-\mu(d-x)}) = e^{-\mu d} \quad [22-3]$$

where d is the total path length through the patient, x is the distance one photon must travel to escape, and $(d-x)$ is the distance the other must travel to escape (Fig. 22-23). Thus, the probability of both escaping the patient without interaction is independent of where on the line the annihilation occurred and is the same as the probability of a single 511-keV photon passing entirely through the patient along the same path. (Equation 22-3 was derived for the case of uniform attenuation, but this principle is valid for nonuniform attenuation as well.)

Even though the attenuation coefficient for 511-keV annihilation photons in soft tissue ($\mu/p = 0.095 \text{ cm}^2/\text{g}$) is lower than those of photons emitted by most radionuclides used in SPECT ($\mu/p = 0.15 \text{ cm}^2/\text{g}$ for 140-keV gamma rays), the average path length for both to escape is much longer. For a 20-cm path in soft tissue, the chance of both annihilation photons of a pair escaping the tissue without

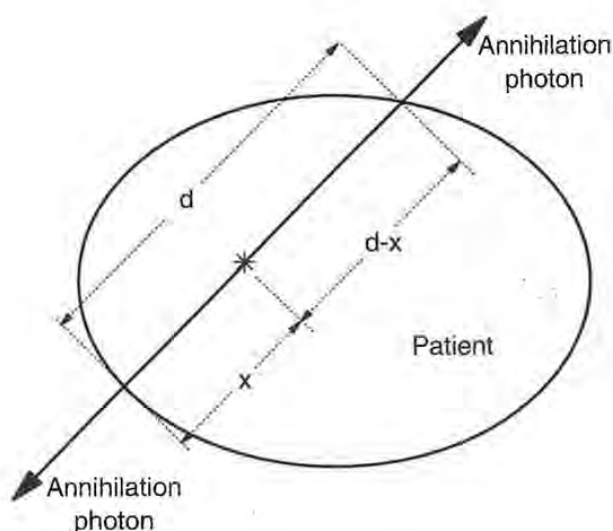


FIGURE 22-23. Attenuation in PET. The probability that both annihilation photons emitted along a particular line of response (LOR) escape interaction in the patient is independent of the position on the LOR where the annihilation occurred.

interaction is only about 15%. Thus, attenuation is more severe in PET than SPECT. The vast majority of the interactions with tissue are Compton scattering. Attenuation causes a loss of information and, because the loss is not the same for all lines of response, causes artifacts in the reconstructed transverse images. The loss of information also contributes to the statistical noise in the images.

As in SPECT, both approximate methods and methods using radioactive sources to measure the attenuation are used for attenuation correction in PET. Most of the approximate methods use a profile of the patient and assume a uniform attenuation coefficient within the profile.

Some PET systems provide one or more retractable positron-emitting sources inside the detector ring between the detectors and the patient to measure the transmission of annihilation photons from the sources through the patient. As mentioned above, the probability that a single annihilation photon from a source will pass through the patient without interaction is the same as the probability that both photons from an annihilation in the patient traveling along the same path through the patient will escape interaction. These sources are usually configured as rods and are parallel to the axis of the scanner (Fig. 22-24). The sources revolve around the patient so that attenuation is measured along all lines of response through the patient. These sources usually contain Ge 68, which has a half-life of 288 days and decays to Ga 68, which primarily decays by positron emission. Alternatively, a source containing a gamma-ray-emitting radionuclide, such as Cs 137 (662-keV gamma ray), can be used for attenuation measurements instead of a positron-emitting radionuclide source. When a gamma-ray-emitting radionuclide is used to measure the attenuation, the known position of the source and the location where a gamma ray is detected together determine a line of response for that interaction. In either case, the transmission data acquired using the external sources are used to correct the emission data for attenuation prior to transverse image reconstruction.

As in SPECT, the transmission data can be obtained before the PET radiopharmaceutical is administered to the patient, or can be obtained during emission

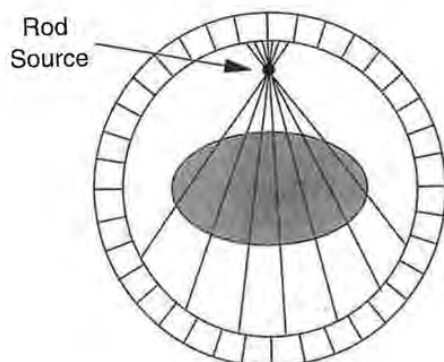


FIGURE 22-24. Rod source for attenuation correction in a dedicated PET system.

imaging. Obtaining the transmission data before the PET radiopharmaceutical is administered to the patient obviates the problem of distinguishing photons emitted by transmission source from those emitted by the radiopharmaceutical. On the other hand, if the transmission data is acquired well before the emission data, it may be difficult to adequately align the two data sets in space. For this reason, the transmission data is often acquired during the acquisition of the emission data.

There are limitations and disadvantages to attenuation correction from transmission measurements. Attenuation correction cannot compensate for the increased statistical noise caused by attenuation. Furthermore, attenuation correction increases the imaging time, although only slightly if performed during emission image acquisition; slightly increases the radiation dose to the patient; and increases the statistical noise in the images. Also, spatial misregistration between the attenuation map and the emission data, most often caused by patient motion, can cause significant artifacts. Attenuation correction may reduce the sensitivity of some clinical studies. Attenuation correction may not be necessary for all PET procedures and, for some procedures in which it is performed, the interpreting physician should review the uncorrected reconstructed images as well.

Alternatives to Dedicated PET Systems—SPECT with High-Energy Collimators and Multihead SPECT Cameras with Coincidence Detection Capability

Some nuclear medicine departments may wish to image positron-emitting radiopharmaceuticals, but lack sufficient demand to justify a dedicated PET scanner. Positron-emitting radiopharmaceuticals also can be imaged using standard scintillation cameras with ultra-high-energy collimators or multihead SPECT cameras with coincidence detection capability.

Some scintillation cameras are capable of performing planar imaging and SPECT of positron emitters using collimators designed for 511-keV photons. These collimators have very thick septa and have very low collimator efficiencies. Despite the thick septa, the images produced suffer from considerable septal penetration. Also, the thick septa produce collimator pattern artifacts in the images. However, attenuation by the patient is much less significant than in coincidence imaging. The gantries of such cameras must be able to support very heavy camera heads, because of the weight of

511-keV collimators and because the camera heads must contain adequate shielding for 511-keV photons from outside the field of view.

Double- and triple-head SPECT cameras with coincidence detection capability are commercially available. These cameras have planar NaI(Tl) crystals coupled to PMTs. With standard nuclear medicine collimators mounted, they function as standard scintillation cameras and can perform planar nuclear medicine imaging and SPECT. With the collimators removed or replaced by axial collimators, they can perform coincidence imaging of positron-emitting radiopharmaceuticals.

When used for coincidence imaging, these cameras provide spatial resolution similar to that of a dedicated PET system. However, in comparison to a dedicated PET system, they detect a much lower number of true coincidences. The smaller attenuation coefficient of sodium iodide and the limited crystal thickness limits the intrinsic efficiency of each detector. For example, the photopeak efficiency of a 0.95 cm ($\frac{3}{8}$ inch) thick NaI(Tl) crystal for annihilation photons is about 12% (see Chapter 21, Fig. 21-9). The fraction of annihilation photon pairs incident on a pair of detectors that yields true coincidences is equal to the square of the photopeak efficiency (assuming that only photopeak-photopeak coincidences are registered). Thus, for 0.95-cm thick crystals, only about 1.5% of annihilation photon pairs reaching the detectors are registered as true coincidences. For this reason, many such cameras have thicker NaI crystals than standard scintillation cameras, with thicknesses ranging from 1.27 to 2.5 cm ($\frac{1}{2}$ to 1 inch). Even modest increases in crystal thickness significantly improve the number of true coincidences detected. However, even with thicker crystals, the coincidence detection sensitivities of multihead scintillation cameras with coincidence circuitry are much lower than those of dedicated PET systems. The lower coincidence detection sensitivity causes more statistical noise in the images than in images from a dedicated PET system. Furthermore, the low coincidence detection sensitivity forces the use of measures, such as wide spacing of axial septa in two-dimensional acquisition or three-dimensional acquisition and wide energy discrimination windows, to obtain adequate numbers of true coincidences. These measures increase the scatter coincidence fraction, producing images of lower contrast than a dedicated PET system. Both the greater noise due to fewer coincidences and the reduced contrast due to the higher scatter coincidence fraction limit the ability to detect small lesions in the patient, in comparison to a dedicated PET system. When these cameras are used with collimators as standard scintillation cameras, their thicker crystals produce slightly worse intrinsic spatial resolution for lower energy photons, such as those emitted by Tc 99m and Tl 201, but have little effect on the system spatial resolution.

Some manufacturers provide two-dimensional data acquisition, whereas others provide three-dimensional acquisition (described earlier in this chapter). The cameras using two-dimensional data acquisition are provided with axial collimators to reduce the singles rates from out-of-slice activity and to reduce the fraction of scatter coincidences.

Because these cameras are either uncollimated or used with axial collimators when performing coincidence imaging, they are subject to very high interaction rates and must have very fast electronics. Because such a system has only two or three detectors (scintillation camera heads) instead of the hundreds of detectors in a dedicated PET system, its performance suffers more from dead-time count losses.

Although many interactions from lower energy photons are rejected by the energy discrimination circuits of the camera, they still cause dead-time count losses and can interfere with the determination of the energies and positions of interactions.

Some cameras are equipped with filters, between the crystal and the axial collimator in two-dimensional mode and in place of the collimator in three-dimensional mode, to preferentially absorb low-energy photons. These filters typically have several layers, such as lead, tin, and copper, with the atomic numbers of the layers decreasing toward the crystal to attenuate characteristic x-rays from the higher atomic number layers.

Clinical experience is accumulating regarding the utility of imaging F-18 FDG using scintillation cameras with either 511-keV collimators or coincidence detection capability. It appears that SPECT with high-energy collimators is acceptable for the determination of myocardial viability, although less accurate than a dedicated PET system. However, attenuation artifacts pose a problem in determining myocardial viability by coincidence imaging with a scintillation camera and accurate attenuation correction may be necessary for this application.

F-18 FDG imaging with high-energy collimators does not appear to be acceptable for brain imaging or for the evaluation and staging of neoplasms, because of its limited efficiency and spatial resolution. On the other hand, coincidence imaging with a scintillation camera may be useful for the detection and staging of neoplasms, but is less sensitive than a dedicated PET system, particularly for detecting smaller tumors. The images from coincidence imaging with a scintillation camera suffer from more statistical noise and have less contrast than those from a dedicated PET system. Coincidence imaging with scintillation cameras for oncologic applications is likely to be more effective in imaging the head, neck, and thorax, where the effects of attenuation and scattering are less severe than in the abdomen and pelvis.

Economical Dedicated PET Systems

Several vendors have developed dedicated PET systems that are less expensive than PET systems with full rings of BGO detectors, but that provide much better coincidence detection sensitivity than double-head scintillation camera coincidence systems. One such design consists of six stationary uncollimated scintillation cameras, each with a curved 2.5 cm thick NaI(Tl) crystal, surrounding the patient (Fig. 22-25, left).

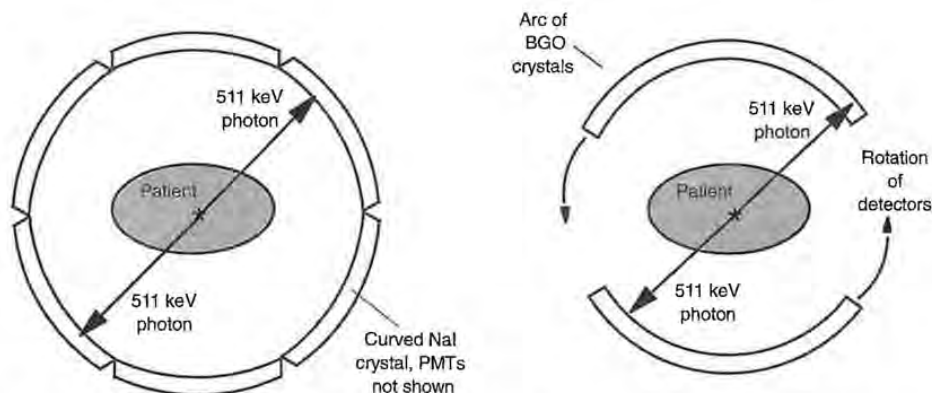


FIGURE 22-25. Less expensive dedicated PET systems. **Left:** This system consists of six scintillation cameras, each with a curved 2.5 cm thick NaI(Tl) crystal, facing the patient. The camera heads are stationary. **Right:** This system consists of two opposed arcs of BGO detectors, of the type shown in Fig. 22-14, facing the patient. To obtain projection data from all lines of response, the detectors revolve around the patient.

A system provided by another vendor consists of two opposed arcs of BGO detectors (Fig. 22-25, right). To obtain projection data over a full 180 degrees around the patient, the detectors revolve about the patient. Both systems use coincidence detection in lieu of collimation to obtain projection data.

Comparison of SPECT and PET

In single photon emission imaging, the spatial resolution and the detection efficiency are primarily determined by the collimator. Both are ultimately limited by the compromise between collimator efficiency and collimator spatial resolution that is a consequence of collimated image formation. It is the use of annihilation coincidence detection instead of collimation that makes the PET scanner much more efficient than the scintillation camera and also yields its superior spatial resolution.

In systems that use collimation to form images, the spatial resolution rapidly deteriorates with distance from the face of the imaging device. This causes the spatial resolution to deteriorate from the edge to the center in transverse SPECT images. In contrast, the spatial resolution in a transverse PET image is best in the center. Table 22-4 compares SPECT and PET systems.

TABLE 22-4. COMPARISON OF SPECT AND PET

	SPECT	PET
Principle of projection data collection	Collimation	Annihilation coincidence detection
Transverse image reconstruction	Filtered backprojection or iterative methods	Filtered backprojection or iterative methods
Radionuclides	Any emitting x-rays, gamma rays, or annihilation photons; optimal performance for photon energies of 100 to 200 keV.	Positron emitters only
Spatial resolution	Depends on collimator and camera orbit Within a transaxial image, the resolution in the radial direction is relatively uniform, but the tangential resolution is degraded toward the center Typically about 10-mm FWHM at center for a 30-cm diameter orbit and Tc-99m (NEMA protocol) Larger camera orbits produce worse resolution	Relatively constant across transaxial image, best at center Typically 4.5 to 5 mm FWHM at center (NEMA protocol)
Attenuation	Attenuation less severe Attenuation correction sources available, but utility not yet established	Attenuation more severe However, accurate attenuation correction is possible, with a transmission source
Cost	Typically \$500,000 for a double-head, variable-angle camera	\$1,000,000 to \$2,000,000

FWHM, full width at half maximum.

Factors Affecting the Availability Of PET

The main factors limiting the availability of PET are the relatively high cost of a dedicated PET scanner and, in many areas of the country, the lack of local sources of F 18 FDG. However, cyclotron-equipped commercial radiopharmacies are now located in many major metropolitan areas. Multihead SPECT cameras with coincidence circuitry and SPECT cameras with high-energy collimators provide less expensive, although less accurate, alternatives for imaging FDG. PET imaging using FDG has become routine at many major medical centers.

Combined PET/X-Ray CT and SPECT/Coincidence Imaging/X-Ray CT Systems

PET imaging, particularly in oncologic applications, often reveals suspicious lesions, but provides little information regarding their exact locations in the organs of the patients. In such situations, co-registration of the PET images with images from another modality, such as x-ray CT or magnetic resonance imaging (MRI), that provides good depiction of anatomy, can be very useful. However, this usually requires moving the patient from the PET system to the CT or MRI system and, once the patient has been moved, it is very difficult to accurately align the PET data with the CT or MRI data. To solve this problem, systems have been developed that incorporate a PET system and a conventional x-ray CT system in a single gantry. The patient lies on a bed that passes through the bores of both systems. A series of CT images is acquired over the same section of the patient as the PET scan, either before or after the PET image acquisition. Because the bed moves in the axial direction, but the patient lies still on the bed, each CT image slice corresponds to a transverse PET image, permitting accurate co-registration. Furthermore, the x-ray CT information can be used to provide attenuation correction of the PET information. In the co-registered images, the PET information is usually superimposed in color on gray-scale CT images.

One manufacturer provides a gantry containing two scintillation camera heads and a simple x-ray CT machine. The scintillation camera heads can be used for SPECT imaging of conventional radiopharmaceuticals and, when placed in the 180-degree orientation, for coincidence imaging of positron-emitting radiopharmaceuticals. The x-ray CT device provides CT image data for image co-registration and for attenuation correction.

SUGGESTED READING

- Bax JJ, Wijns W. Editorial: Fluorodeoxyglucose imaging to assess myocardial viability: PET, SPECT, or gamma camera coincidence imaging. *J Nucl Med* 1999;40:1893–1895.
- Cherry SR, Phelps ME. Positron emission tomography: methods and instrumentation. In: Sandler MP, et al., eds. *Diagnostic nuclear medicine*, 3rd ed., vol. 1. Baltimore: Williams & Wilkins, 1996:139–159.
- Gelfand MJ, Thomas SR. *Effective use of computers in nuclear medicine*. New York: McGraw-Hill, 1988.
- Groch MW, Erwin WD. SPECT in the Year 2000: Basic principles. *J Nuclear Med Tech* 2000;28:233–244.
- Groch MW, Erwin WD, Bieszk JA. Single photon emission computed tomography. In: Treves ST, ed. *Pediatric nuclear medicine*, 2nd ed. New York: Springer-Verlag, 1995:33–87.

- Madsen MT. The AAPM/RSNA physics tutorial for residents. Introduction to emission CT. *RadioGraphics* 1995;15:975-991.
- Miller TR. The AAPM/RSNA physics tutorial for residents. Clinical aspects of emission tomography. *RadioGraphics* 1996;16:661-668.
- NEMA. Recommendations for implementing SPECT instrumentation quality control. *J Nucl Med* 2000;41:383-389.
- Votaw JR. The AAPM/RSNA physics tutorial for residents. Physics of PET. *RadioGraphics* 1995;15:1179-1190.
- Yester MV, Graham LS, eds. *Advances in nuclear medicine: The medical physicist's perspective*. Proceedings of the 1998 Nuclear Medicine Mini Summer School, American Association of Physicists in Medicine, June 21-23, 1998, Madison, WI.