

Chapter 1

Introduction and Summary

This document is latex formatted chapters of the book

Mathematics and Physics of Emerging Biomedical Imaging
Copyright 1996 by the National Academy of Sciences.

The first introductory chapter, complete with many pictures, is available on-line at:

<http://www.nap.edu/readingroom/books/biomedical/>

The book was apparently written by a committee:

COMMITTEE ON THE MATHEMATICS AND PHYSICS OF EMERGING DYNAMIC BIOMEDICAL IMAGING

Thomas Budinger, Lawrence Berkeley National Laboratory, Co-chair
Felix Wehrli, University of Pennsylvania Medical Center, Co-chair
S. Morris Blumenfeld, General Electric Medical Systems
F. Alberto Grunbaum, University of California at Berkeley
R. Mark Henkelman, University of Toronto
Paul C. Lauterbur, University of Illinois at Urbana-Champaign
Wilfried Loeffler, Siemens Medical Systems, Inc.
Frank Mattereder, University of Muenster
Sarah Jane Nelson, University of California at San Francisco
Lawrence Shepp, AT&T Bell Laboratories
Robert G. Shulman, Yale University
Benjamin Ming Wah Tsui, University of North Carolina at Chapel Hill

Scott T. Weidman, Senior Staff Officer, Board on Mathematical Sciences
Robert L. Riemer, Associate Director, Board on Physics and Astronomy
Constance M. Pechura, Director, Board on Biobehavioral Sciences and Mental Disorders

1.1 PREFACE

The Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging was constituted in 1993 and given the charge to "write a report that gives a survey of the emerging contributions of the mathematical sciences and physics to dynamic biomedical imaging and identifies and recommends specific mathematical sciences and physics research to accelerate the development and implementation of new medical imaging systems." At its first meeting, the committee discussed the frontiers of biomedical imaging that could profit from more involvement from physicists and mathematical scientists, outlined its proposed report, and identified individuals, listed below, who could supplement the committee's expertise in documenting these frontiers and the related research opportunities. At its subsequent two meetings, the committee drew on the large quantity of valuable drafts to generate the survey it envisioned. It is hoped that the present report will provide a readable introduction to emerging techniques of biomedical imaging for mathematical scientists and physicists and encourage some of them to apply their skills to the research challenges that will make these emerging techniques practical.

The committee gratefully acknowledges the substantial contributions of the following people, who provided material for the committee to incorporate in its report:

Richard Albanese, Brooks Air Force Base
Robert Alfano, City University of New York
Simon Arridge, University College, London
Randall Barbour, SUNY Health Science Center at Brooklyn
Harrison Barrett, University of Arizona
James Berryman, Lawrence Livermore National Laboratory
Douglas P. Boyd, IMATRON-West
Britton Chance, University of Pennsylvania
Margaret Cheney, Rensselaer Polytechnic Institute
Rolf Clack, University of Utah
James G. Colsher, GE Medical Systems
Robert Cox, Medical College of Wisconsin
Michel Defrise, Vrije Universiteit
Charles L. Dumoulin, General Electric R&D Center
Alan C. Evans, Montreal Neurological Institute
Stuart Foster, University of Toronto
C. Franzone, University of Pavia
E.C. Frey, University of North Carolina at Chapel Hill
Michael M. Graham, University of Washington
Enrico Gratton, University of Illinois at Urbana-Champaign
Peter J. Green, University of Bristol
James F. Greenleaf, Mayo Clinic
Grant T. Gullberg, University of Utah
Semion Gutman, University of North Carolina at Charlotte
E. Mark Haacke, Mallinckrodt Institute of Radiology
Dennis M. Healy, Dartmouth College
Manfried Hoke, University of Muenster
Paul W. Hughett, Lawrence Berkeley National Laboratory
James Hyde, Medical College of Wisconsin
V. Isakov, Wichita State University
Steven A. Johnson, University of Utah
Valen E. Johnson, Duke University
Willi A. Kalender, University of Erlangen-Nuremberg
Linda Kaufman, AT&T Bell Laboratories
Ronald Kikinis, Harvard Medical School
Michael A. King, University of Massachusetts Medical School
Michael Klibanov, University of North Carolina at Charlotte
David Levin, University of Chicago
Tom Lewellen, University of Washington
Jorge Llacer, Lawrence Berkeley National Laboratory
Bernd Luetkenhoener, University of Muenster
Albert Macovski, Stanford University
Ravi S. Menon, University of Western Ontario at London
Michael I. Miller, Washington University
Charles Mistretta, University of Wisconsin at Madison
Adrian Nachman, University of Rochester
Claude Nahmias, Chedoke-McMaster Hospitals
William D. O'Brien, Jr., University of Illinois at Urbana-Champaign
Matthew O'Donnell, University of Michigan
John Ollinger, Washington University
Arnulf Oppelt, Siemens Medical Engineering Group
Walter W. Peppler, University of Wisconsin
Stephen M. Pizer, University of North Carolina at Chapel Hill
Jack Reid, Drexel University
Joel G. Rogers, TRIUMF, University of British Columbia
Yoram Rudy, Case Western Reserve University
David Saloner, Veterans Affairs Medical Center, San Francisco
Guenter Schwierz, Siemens Medical Engineering Group
V. Sharafutdinov, Institute of Mathematics, Novosibirsk
Gunnar Sparr, Lund Institute of Technology
Terry Spinks, Hammersmith Hospital
J. Sylvester, University of Washington
Robert Turner, University of London
Eugene Veklerov, Lawrence Berkeley Laboratory
Robert Weisskoff, Massachusetts General Hospital

The committee is also grateful to the six anonymous reviewers, whose comments strengthened this report considerably.

Chapter 2

X-Ray Projection Imaging

2.1 Introduction

X-ray projection imaging—the familiar “x-ray”—has been and continues to be the most widely used format for diagnostic imaging in spite of important and exciting advances made, for example, in the areas of computed tomography and magnetic resonance imaging.

X-ray projection imaging dates back to Röntgen’s discovery of the x-ray in 1895. In subsequent decades techniques were developed involving the use of contrast materials to enhance the images of blood vessels, gastrointestinal structures, and other anatomy. In 1930 the technique of film subtraction angiography was developed and, in spite of the inconvenient delay required to complete the film processing, enjoyed widespread use for many decades.¹

Before the invention of the image intensifier in 1950, x-ray diagnosis was accomplished using either static film images or fluoroscopy, which involved direct viewing of fluorescent screens after the viewer adapted to the dark. The invention of the image intensifier was a major advance not only because it improved fluoroscopy but also because its introduction of an electronic image signal facilitated the development of real-time digital imaging techniques such as digital subtraction angiography (DSA), the digitized analog of film subtraction angiography.

At about the same time that DSA was introduced, digital projection images were obtained with computed tomography (CT) systems and dedicated line scan systems. Line scan images,² which are used routinely to provide localization images prior to CT, benefit from the low scatter associated with the linear collimation employed.

The emergence of digital (as opposed to film) techniques for x-ray projection imaging, combined with the fact that several other diagnostic modalities (e.g., nuclear medicine, CT, and more recently ultrasound and magnetic resonance imaging) provide digital data, has led to increased efforts focused on the possibility of completely digital radiology departments. The movement toward this has been slowed by the complexity of the task and also by the lack of a universally acceptable digital detector for x-ray projection images. While image-intensifier television systems have been adequate for digital fluoroscopy applications, a substitute for film has not yet been found that provides comparable diagnostic accuracy over a wide range of applications.

The development of photo-stimulable phosphor plates in the mid 1980s provided a detector with several attractive properties, including a digital format and greatly increased dynamic range. This modality compares well with film in all but a few applications and appears superior for bedside radiography because of its improved dynamic range. However, it does not have sufficient resolution for mammography and has the disadvantage, shared by film, that cassettes must be handled and processed.

The availability of rapidly obtained digital images has led to investigations of dual energy techniques for chest radiography and cardiac imaging.³ Although most of these investigations have been carried out using conventional x-ray sources, there are ongoing investigations of alternative sources for producing narrow band radiation for applications of this type.

While it is recognized that film is unique in its ability to serve as detector, display mechanism, and archival storage medium, the availability of digital techniques offers the possibility of separately optimizing these functions. Significant current research activity revolves around the development and optimization of detectors, sources, and

¹ Film subtraction angiography is a technique that uses positive and negative film images taken before and after the injection of contrast agents to produce subtraction images showing only the vascular structures affected by the contrast injection, without superposition of other structures (e.g., bones).

² Produced by sequentially scanning a linear (one line of data at a time) detector over the area of interest.

³ Dual-energy techniques involve acquiring two images from the same area using different x-ray spectra (by changing the tube voltage) in order to exploit differences in the attenuation coefficient of different body tissues as a function of the x-ray energy.

related imaging apparatus and procedures for several imaging applications. The remainder of this chapter summarizes these challenges within the context of some of the most significant areas of clinical applications of projection imaging.

2.2 Mammography

In mammography the development of digital techniques is based primarily on the desire to provide a lower-dose examination for improved detection and perhaps characterization of lesions. Because of the desire to image micro-calcifications, system resolution must be on the order of 50 microns. Additionally, in order to achieve the required resolution of low-contrast differences inherent in mammography, the detector must be capable of 12-bit contrast resolution.

Although digital detectors already exist for imaging a part of the breast, primarily for stereotaxic localization,⁴ efforts are being made to develop detectors capable of imaging the entire breast. These developments can be divided into systems that perform some form of scan to acquire the images versus area detectors that each simultaneously acquire an image of the entire breast.

2.2.1 Scanning Methods

In some ways the technical requirements for a scanning system using digital detectors are lower than the requirements for area systems because the number of discrete detection elements can be significantly reduced. However, the mechanism to produce the scanning motion can add significant complexity to the system.

In addition to having lower scatter, scanned systems offer the potential for image equalization through spatially varying the exposure. This possibility, previously explored for chest radiography, has recently been investigated in a film system. The film density is equalized throughout the mammogram using a raster scanned x-ray beam with feedback from an opposing detector. This approach offers the advantage that higher-contrast film, which can better display the small structures of the breast, can be used because the large intensity variations through the breast have been reduced. In the case of digital detectors, which enable manipulations of the contrast after image acquisition, scanning equalization still offers the potential advantage of a spatially uniform signal-to-noise ratio and more efficient dose utilization. Just as was done in chest radiography, the mammographic equalization system presumably could be implemented in a line-scanned configuration to reduce tube loading and imaging time.

There are several slot or multislot scanning systems reported or under development.⁵ Detector schemes for such systems include phosphor fiber optically coupled to charge coupled devices (CCDs), systems utilizing direct detection of x-rays by glass or plastic scintillation fibers coupled to CCDs, or silicon photodiode arrays.

2.2.2 Area Detectors

Photostimulable phosphor plate detectors have met with some success in chest radiography but have had relatively little acceptance in mammography, presumably owing largely to the limited spatial resolution. In addition, they have reduced detector quantum efficiencies (DQEs)⁶ relative to those of film/screen systems. There is ongoing activity to improve both aspects of the plate performance, and additional phosphor materials are being investigated. In addition, the expiration in the next year or two of some of the patents on this technology should increase competitiveness and accelerate developments in this area. Although much has been borrowed from conventional screen technology, it is likely that significant improvements are still possible.

Several small-field-of-view (5 cm × 5 cm) commercial digital mammography systems are available and are used primarily for stereotactic needle biopsy of the breast. These systems all employ detection of x-rays by fluorescent screens coupled to CCD detectors by a lens or fiber-optic line. With future improvements in CCD technology, which are also being driven by other industries, it is likely that either single or multiple CCD detectors capable of imaging the entire breast will be possible.

Amorphous silicon sheets with multiple discrete detectors are under development. The silicon photodetector is usually coupled to a scintillation screen to convert the x-ray energy to visible light that is then detected by the silicon

⁴Stereotaxic localization is the three-dimensional localization of lesions for biopsies or surgery using imaging techniques, usually requiring the fixation of the organ under investigation within a mechanical frame during imaging and surgery.

⁵Slot scanning imaging systems are similar to line scan imaging and use linear detector systems to sequentially scan the imaging field.

⁶The quantum efficiency is a measure of detector sensitivity related to its ability to interact with a certain fraction of the incoming x-ray or γ -quanta. The quantum efficiency is related to the linear attenuation coefficient. Often geometric efficiencies (active versus inactive area of the detector) are included.

array. Such arrays have insufficient spatial resolution for mammography at this time, but continued development may make better resolution possible.

Selenium plates were used previously as an x-ray detector in xeromammography.⁷ Interest in selenium as a detector has been rekindled with the development of some new readout strategies. One of these involves the use of integral discrete detectors on a glass substrate. Another uses a luminescent liquid toner that is scanned by a laser, producing luminescence that is recorded by a charged selenium photodetector plate. One advantage of selenium detectors is that they have inherently high resolution. Whereas previous readout schemes had limited resolution due to the discrete readout process, in this case the luminescent process would be limited by the laser spot size, which can be made very small. Another advantage of selenium is that it can be made quite thick in order to detect a large fraction of the incident radiation without sacrificing resolution. Therefore, the DQE of selenium detectors can be quite good.

Another area of interest is in the development of phosphors that utilize some mechanism to contain the lateral spread of visible light as it exits the detector. Two separate, fairly similar efforts address this objective. The first uses doped glass fibers to produce fluorescent light that stays in the fiber, as in an optical fiber. The other effort involves the use of plastic fibers activated by organic fluorescent dyes to a certain depth at the end of the fiber. In both cases the fiber axis is oriented parallel to the direction of the x-rays. In this way the effective phosphor thickness can be large without sacrificing resolution.

Limitations on intrinsic detector resolution may be relaxed if magnification techniques can be used, although this approach is usually limited by the loss of resolution due to focal spot blurring. The use of x-ray capillary optics is being investigated as a means for achieving magnifications of up to 2.0 without additional focal spot blurring. The capillary optics, which channel the radiation detected at the opposing side of the patient to conventional phosphor plates, prevent focal spot blurring beyond that which exists at the exit plane. Another possible way to achieve large magnifications is through the use of high-output, small-focal-spot x-ray sources, exploration of which has already begun.

2.3 Chest Radiography

2.3.1 Scanning Methods

For chest radiography several slot or area scanning systems employing image equalization have been previously investigated. A scanning slit system capable of producing dual-energy images has also been reported. The narrow slit utilized provided for excellent images but led to high tube loading. It may be possible to increase the slit width to a more favorable size and still maintain image quality. However, these systems have met with limited success.

2.3.2 Area Detectors

Recently there has been significant improvement in conventional film/screen cassettes. The new cassettes utilize asymmetric front and back screens, different front and rear film emulsions, and anticrossover technology so that light from the front screen is not detected by the back emulsion and vice versa. The advantage is that the resolution, contrast, and speed of the front film/screen combination can differ from that of the back combination, permitting variations in image properties throughout the image. This technology has been applied specifically in chest radiography to yield high resolution and contrast in the lung regions and increased film density, good contrast, and somewhat lower resolution in the mediastinum, behind the heart, and below the diaphragm.

Amorphous silicon and selenium detectors (described in the section above on mammographic detectors) are also being investigated for chest radiography. These detectors may have a greater likelihood of success in the chest application, since the demands for resolution are lower for chest radiography than for mammography.

A digital chest radiography system utilizing a selenium detector has recently been developed. The detector is a drum of amorphous selenium. After exposing the selenium, an array of sensitive electrometer probes reads the charge pattern as the drum is rotated at high speed. The dynamic range of the detector is very good and the DQE is reported to be better than that for either phosphor plates or conventional film/screen combinations.

Photostimulable phosphor plate systems are continuing to gain acceptance for bedside chest radiography but are not widely used for dedicated chest radiography, which has much higher requirements for image quality. However, continued improvements in phosphor plate technology may eventually increase its broader acceptance.

⁷Xeromammography is a filmless method of breast imaging that uses a xerographic technique for image generation: x-rays produce an electrical charge distribution that is reproduced on paper through the familiar photocopying or laser printing process.

2.4 Digital Fluoroscopy

Digital fluoroscopy includes image-intensifier television camera systems such as those used for DSA, non subtractive cardiac imaging, and gastrointestinal fluoroscopy. Developments in this area include the use of higher-resolution video cameras, potentially higher resolution detectors, image acquisition strategies to permit imaging of moving tables or multiple angles during an angiographic injection, and lower-lag CCD camera systems for cardiac applications.

Compensation for bright spots in the detected x-ray field by means of a selectable configuration of a basic series of spatially variable filter wheels has been investigated and holds promise for solving one of the most important problems in the imaging of low-contrast objects.

It has been known for some time that the use of x-ray contrast materials with k-edges higher than those of iodine and barium would provide greatly improved imaging performance because of the favorable trade-off between contrast loss and increased tissue transmission. Recent work has confirmed the potential advantages of contrast materials based on gadolinium. There is also continued development of iodinated contrast materials. The development of agents with lower osmolality appears to be more significant in terms of patient tolerance than for any advantages in x-ray attenuation. However, these agents may produce fewer misregistration artifacts in subtraction imaging. Agents designed to remain for longer times in the vascular space may be of benefit for quantitative applications where leakage into interstitial space can complicate interpretation of contrast time curves.

Dual-energy cardiac DSA systems using conventional x-ray sources in conjunction with energy switching and rotating filter wheels are currently under investigation and have been shown to be effective in removing tissue-related artifacts that impair various quantitation tasks involving injected iodine.

Specialized sources such as synchrotron radiation, characteristic radiation, channeling radiation, parametric radiation, and laser sources are under consideration as possible sources for dual-energy DSA. The application most often considered is intravenous coronary angiography, although synchrotron radiation has also been applied to achieve high-resolution three-dimensional microscopy of trabecular bone. Due to the large x-ray doses required with these sources, screening of asymptomatic patients is not justifiable, and the economics of the more expensive of these alternatives would be difficult to rationalize. Additionally, the x-ray approach to coronary angiography, which has progressed slowly in the last decade, should be weighed against recent advances in magnetic resonance imaging of coronary arteries (see Figure 1.3).

2.5 Portal Imaging

Electronic portal imaging devices for verification of radiotherapy treatment fields are rapidly being developed.⁸ These systems offer the advantage, in some cases, of almost real-time operation so that field alignment errors can be corrected quickly.

Several technologies are under investigation. The most common systems use metal plates combined with phosphor materials that are viewed by a video camera. The early systems used a mirror to feed images to the video camera, whereas current systems employ fiber-optic coupling. Amorphous silicon arrays are being evaluated to replace the camera in a direct digital system.

Scanning liquid ionization chambers are also being developed. These devices are similar to conventional ionization chambers except that in place of air they use a liquid. This substitution offers increased detection efficiency in proportion to the ratio of liquid and air densities. Another system under development uses a scanned linear array of high-voltage rectifier diodes as a direct solid-state detector. Other linear scanning array detectors are being developed that use CdTe or zinc tungstate crystals coupled to a photodiode.

2.6 Research Opportunities

- Development of electronic planar array detectors with adequate resolution, size, reliability, and quantum efficiency.
- Development of digital display systems of sufficient resolution and dynamic range.
- Development of means to detect and use the information in scattered radiation, including mathematical correction schemes.

⁸These are filmless real-time imaging systems with large apertures that verify the treatment field by employing a geometry similar to the geometry of the therapy equipment.

- Re-evaluation by physicists of the prospects for multiple monochromatic sources, novel x-ray generation techniques, and holographic methods, ambitious research directions that have been resistant to progress.

2.7 Suggested Reading

1. Antonuk, L.E., Yorkston, J., Huang, W., Boudry, J., and Morton, E.J., Large area, flat-panel a-Si:H arrays for x-ray imaging, *Proc. SPIE* **1896** (1993), 18–29.
2. Barnes, G.T., Wu, X., and Sanders, P.C., Scanning slit chest radiography: A practical and efficient scatter control design, *Radiology* **190** (1994), 525–528.
3. Boyer, A.L., Antonuk, L., Fenster, A., Van Herk, M., Meertens, H., Munro, P., Reinstein, L.E., and Wong, J., A review of electronic portal imaging devices (EPIDs), *Med. Phys.* **19** (1992), 1–16.
4. Endorf, R.J., Kulatunga, S., Spelic, D.C., DiBianca, F.A., and Zeman, H.D., Development of a dual-energy kinesthetic charge detector, *Proc. SPIE* **1896** (1993), 180–191.
5. Gray, J.E., Stears, J.G., Swenson, S.J., and Bunch, P.C., Evaluation of resolution and sensitometric characteristics of an asymmetric screen-film imaging system, *Radiology* **188** (1993), 537–539.
6. Kinney, J.H., Lane, N.E., and Haupt, D.L., In vivo three-dimensional microscopy of trabecular bone, *J. Bone Miner. Res.* **10** (1995), 264–270.
7. Maidment, A.D., Fahrig, R., and Yaffe, M.J., Dynamic range requirements in digital mammography, *Med. Phys.* **20** (1993), 1621–1633.
8. Maidment, A.D.A., Yaffe, M.J., Plewes, D.B., Mawdsley, G.E., Soutar, I.C., and Starkoski, B.G., Imaging performance of a prototype scanned-slot digital mammography system, *Proc. SPIE* **1896** (1993), 93–103.
9. Morris, T., X-ray contrast media: Where are we now, and where are we going?, *Radiology* **188** (1993), 11–16.
10. Plewes, D.B., and Wandtke, J.C., A scanning equalization system for improved chest radiography, *Radiology* **142** (1982), 765–768.
11. Sabol, J., Soutar, I., and Plewes, D., Mammographic scanning equalization radiography, *Med. Phys.* **20** (1993), 1505–1515.
12. van Lysel, M., Optimization of beam parameters for dual energy digital subtraction angiography, *Med. Phys.* **21** (1994), 219–226.
13. Vasbloom, H., and Schultze Kool, L.J., AMBER: A scanning multiple-beam equalization system for chest radiography, *Radiology* **169** (1988), 29–34.
14. Wandtke, J.C., Bedside chest radiography, *Radiology* **190** (1994), 1–10.

Chapter 3

X-Ray Computed Tomography

3.1 Introduction

3.1.1 History

Computed tomography (CT) was the first non-invasive radiological method allowing the generation of tomographic images of every part of the human body without superimposition of adjacent structures. Even more importantly, a contrast resolution could be achieved that for the first time in radiology permitted the differentiation of soft tissue inside the highly attenuating skull. The technological improvements in the first 20 years of CT can best be characterized by a comparison of typical scan parameters from 1972 and 1994, as shown in Table 3.1.

The greatest progress has been made in reducing scan times and in improving spatial resolution. Further increasing the contrast resolution will require an increase in the patient dose, since the efficiency of the detector systems is already close to its theoretical limits. An improvement in the transaxial or planar spatial resolution is sensible only if the resolution as measured along the scanner axis (i.e., the slice thickness) is improved simultaneously. With today's typical scan parameters there is a discrepancy between the planar resolution, which is less than 1 mm, and the slice thickness, which customarily is several millimeters. With the recent introduction of spiral CT (section 3.3), x-ray CT, like other imaging modalities, has started to move from a slice-by-slice to a volume imaging method, with more isotropic spatial resolution in all three dimensions allowing a better three-dimensional representation of anatomical structures.

3.1.2 Principle of Operation

The attenuation of monochromatic x-rays in homogeneous objects is governed by

$$I = I_0 e^{-ax} , \tag{3.1}$$

where I is the x-ray intensity behind the object, I_0 is the x-ray intensity without the object, x is the length of the x-ray path through the object, and a is the linear attenuation coefficient of the material for the x-ray energy employed.

For inhomogeneous objects like the human body, the attenuation of x-rays consequently can be described by

$$I = I_0 e^{-\int a(x) dx} . \tag{3.2}$$

Table 3.1 Technological Improvements in the First 20 Years of Computed Tomography

Scan Parameter	Typical Values, 1972	Typical Values, 1994
Matrix size	80 × 80	1024 × 1024
Spatial resolution	3 line pairs/cm	15 line pairs/cm
Contrast resolution	5 mm/5 HU/50 mGy	3 mm/3 HU/30 mGy
Scan time	5 min	< 1 sec
Data size per scan	50 kByte	2 MByte
X-ray power	2 kW	40 kW

Definitions: mGy = milligrays, where a gray is the SI unit for radiation dose. HU = Hounsfield units; the contrast scale used in CT is defined by zero attenuation (air) = -1000 HU and attenuation coefficient of water = 0 HU (see equation 3.3).

Figure here Figure 3.1. Scanning principles of computed tomography systems. Left: Fan beam systems employ a multicellular detector system rotating about the patient together with the x-ray tube. Right: Ring detector based systems have a fixed detector ring, operate with a fan-shaped x-ray beam, and require only a rotational movement of the x-ray tube or the x-ray focal spot in the case of electron beam type x-ray generation. (Reprinted with permission from E. Krestel, ed., *Imaging Systems for Medical Diagnosis*, Siemens Aktiengesellschaft, Berlin and Munich, 1990. Copyright Publicis MCD Werbeagentur GmbH (Siemens AG), 1990.)

Figure here Figure 3.2. CT detectors. Left: Scintillation crystal with photo-sensitive semiconductor detector. Right: Rare-gas ionization chamber detector. (Reprinted with permission from E. Krestel, ed., *Imaging Systems for Medical Diagnosis*, Siemens Aktiengesellschaft, Berlin and Munich, 1990. Copyright Publicis MCD Werbeagentur GmbH (Siemens AG), 1990.)

This model is explored further in section 14.1.1.

In CT scanners the x-ray attenuation according to equation 3.2 is measured along a variety of lines within a plane perpendicular to the long axis of the patient with the goal of reconstructing a map of the attenuation coefficients a for this plane. The resulting attenuation coefficients are usually expressed with reference to water:

$$a(\text{in HU}) = 1000 \frac{a_{\text{patient}} - a_{\text{water}}}{a_{\text{water}}}. \quad (3.3)$$

Equations 3.1 and 3.2 ignore the fact that the spectrum of the x-rays entering the patient is generally non-monochromatic. It turns out that the resulting “beam hardening artifacts” can be corrected to a sufficient degree by a simple nonlinear correction to the measured line integrals.¹

Figure 3.1 shows the two principal measurement geometries used in today’s CT scanners. In fan beam systems (Fig. 3.1, left) a multicellular linear detector array and the x-ray tube are mounted on the same carrier, which rotates around the patient. A fan-shaped x-ray beam is utilized to successively measure the x-ray attenuation across the patient for the required number of projection angles. The image then can be reconstructed either after rearranging the central projections into parallel projections with parallel beam reconstruction techniques, or more directly by using the central projections with a modified fan beam reconstruction algorithm. A variant of the fan beam system employs a number of x-ray tubes and detector arrays rotating around the patient with the x-ray focal spots shining through the gaps between the detector arrays. This configuration allows for a shorter exposure time since the system has to be rotated only through a small angle rather than 180° or more in order to obtain a complete data set for image reconstruction. However, because of the very high cost involved and because of the improved speed of slip ring or electron beam based systems, this variant of the conventional fan beam system has not gained wide distribution.

Ring detector based systems (Fig. 3.1, right) use a complete stationary ring of detector elements surrounding the patient port and move only the x-ray tube—or, in the case of electron beam systems, the focal spot—around the patient. The measurement and reconstruction geometry also employs a fan of x-rays through the patient in the same way as with fan beam systems. However, particularly in electron beam based systems, it is often difficult to move the x-ray focal spot in the same plane that is occupied by the ring detector. In this case the focal spot has to follow orbit in a plane adjacent to the detector system, which represents an imperfection in the measurement geometry and can lead to typical artifacts (see also section 3.2.3).

3.2 Present Status of CT Instrumentation and Technology

3.2.1 X-Ray Tubes

The dramatic increase in the performance of conventional—i.e., non-electron-beam based—CT scanners over the past 20 years has required significant improvements in the technology of the x-ray tubes used. Today, typically, a metal (tungsten-rhenium-molybdenum) target is brazed to a thick graphite block for improved heat dissipation. Target diameters of up to 20 cm are used with rotational speeds of up to 10,000 rpm. Typically the target is operated at a temperature of 1200°C and the area around the focal spot reaches over 2500°C. The maximum heat dissipation of these tubes can reach up to 10 kW. Still, the major limitation in the use of CT scanners imposed by the x-ray tube is the maximum permissible sustained x-ray power. A further increase will require major developments in target cooling (e.g., liquid cooled-rotating targets), bearing systems, and generation of the electron beam.

¹Beam hardening is the effect of different attenuations of the different parts of the x-ray spectrum when the x-ray passes through an object.

3.2.2 Detector Systems

Figure 3.2 gives an overview of the two types of detector systems employed with CT. Both types of systems are usually arranged in linear arrays with up to about 1,000 detector elements or as rings with even more elements in the case of systems with a geometry as shown on the right in Figure 3.1. At least one commercial system uses two adjacent rows of detector elements in order to scan two slices simultaneously. For scintillation based detectors the most commonly used scintillator materials have been cesium iodide (CsI) and cadmium tungstate, as well as ceramic scintillators employing rare-earth oxides and oxysulfides. A particularly critical parameter for CT detector scintillators is their afterglow behavior. Since CT data acquisition systems do not use pulse counting as in most nuclear applications, but rather integrate the detector current, the afterglow leads to measurement errors near tissue/air and tissue/bone interfaces. Data correction schemes have been developed to at least compensate for first-order effects related to the detector afterglow. Compared to the rare-gas ionization detector, the scintillation detector usually has a lower geometric efficiency (ratio of active versus ineffective frontal area), which is, however, generally more than compensated for by the latter's higher quantum efficiency.² In fan beam systems a collimator is usually employed in front of the scintillator elements to reduce the effects of scattered x-rays. In gas detector arrays the chamber walls effectively serve the same purpose by acting as a collimator aligned with the position of the x-ray focal spot. This self-collimating effect of gas detectors, on the other hand, excludes their application in ring detector based systems, since the focal spot of the x-ray tube moves relative to the detector elements.

3.2.3 Image Artifacts

Image imperfections or artifacts observed in CT systems can be classified according to their origin:

1. Physics-related causes

- **Quantum noise:** Physically unavoidable statistical fluctuations in the x-ray beam intensity will show up as signal-dependent noise in the projection data. A reduction of the noise level requires an increase in x-ray intensity or data acquisition time, along with an increase in patient dose. The signal dependency of the quantum noise leads to typical radial noise structures in the reconstructed images.
- **X-ray scattering by the patient:** In fan beam type systems it is possible to use detectors that are sufficiently collimating to reduce the effect of scatter adequately. In detector ring based systems, however, mathematical scatter models have to be used to correct for this effect.
- **Beam hardening:** Simple empirical corrections are usually sufficient to compensate for this effect for all diagnostic purposes. In particular situations such as when the amount of bone in a particular slice is very large, residual beam hardening effects (e.g., the dark bands typically seen between the petrosal bones in head scans) may show up.
- **Nonlinear partial volume effects:** As a consequence of the exponential, thus nonlinear, attenuation of radiation when high-contrast structures are immersed only partially in the imaged slice, the measured data become inconsistent with the mathematical model used for reconstruction. Although correction methods have been devised, they have not been widely accepted owing to the associated increase in expense, measurement time, and complexity. The simplest and most effective way to reduce nonlinear partial volume effects is to reduce the slice thickness.

2. Patient-related effects

- **Movement of the patient or of individual organs:** Typical motion artifacts will show up as stripes or groups of stripes and streaks. The intensity will depend on the contrast of the moving structures. Although in the images these artifacts can usually be easily identified as motion related, the combination of several artifacts can make it considerably more difficult to interpret the images. The most effective way to reduce motion artifacts, of course, is to speed up the data acquisition time accordingly.
- **Metallic foreign objects:** Due to the strong attenuation of metal (for example, in dental fillings), the dynamic range of the detector and data acquisition system is often not large enough to accurately accommodate the measurement of the projection data. To make things worse, in such a situation nonlinear partial volume effects usually amplify inconsistencies in the projection data. As a consequence the images show streaking artifacts originating from the metallic objects.

3. System-related causes

²The quantum efficiency is a measure of detector (material) sensitivity related to its ability to interact with a certain fraction of the incoming x-ray or γ -ray quanta. The quantum efficiency is related to the linear attenuation coefficient.

- Insufficient calibration of detector sensitivity
- Time-, temperature-, or irradiation-history-dependent drift in detector sensitivity
- Inadequacies in the reconstruction algorithm
- Non-uniform scanning motion
- Fluctuations in x-ray tube voltage

All of the above errors will show up in the reconstructed images with different typical patterns that vary further with the scanning geometry employed. For example, detector-related problems in fan beam systems usually manifest themselves as ring-shaped artifacts, whereas in systems based on stationary ring detectors they will have the character of streaks or stripes.

3.2.4 Quantitative CT

Although in the overwhelming majority of all examinations the resulting images are inspected only visually and qualitatively by the radiologist providing the diagnosis, a few special applications require absolute evaluation of the attenuation coefficients. These quantitative CT applications have a fairly long tradition, particularly in the case of bone mineral density measurements for diagnosis and therapeutic followup of osteoporosis. Here the measured attenuation coefficients or “Hounsfield numbers” are usually compared with simultaneous measurements made on calibrated, calcium-containing inserts. By using inserts with calcium content similar to that of the vertebra to be investigated, patient-dependent beam hardening effects can be compensated for, at least to a first-order approximation. A further refinement requires measuring attenuation data with two different x-ray spectra by changing the x-ray tube voltage. By using these so-called dual-energy scans it is possible to distinguish the individual contributions of calcium and waterlike tissue to the total measured attenuation coefficient, and thus better determine the bone mineral content of the vertebra.

There are a number of less customary quantitative CT procedures such as measurement of tissue perfusion with dynamic CT or xenon CT.³ Measurement of lung density represented an area of great interest very early on but did not become feasible in practice until spirometric (related to a measurement of the respiratory cycle) control of the scan and suitable automated evaluation methods were available. Apart from documenting findings objectively, as is frequently necessary in occupational diseases, the primary hope is to be able to detect diseases sensitively and early by selective analysis of different areas, e.g., by automated evaluation of the peripheral parts of the lungs.

3.2.5 Requirements for High-Speed CT

Most routine applications of CT ask for the best possible image acquired within the shortest possible time. The requirement for short acquisition times comes both from the necessity to effectively freeze the motion of organs like the heart and from the desire to cover large areas of the patient in the shortest possible time to improve patient comfort and patient throughput. To avoid the relatively greater effects of quantum noise associated with shorter scan times, the most critical and technically most demanding requirement is improvement to both instantaneous and sustained x-ray power from the tube or electron beam system.

3.3 Spiral CT

With the introduction of slip ring based scanners in the mid-1980s, continuous-volume data acquisitions became possible. Such imaging is achieved by moving the patient slowly through the gantry while continuously acquiring data. The resulting helical movement of the x-ray tube focal spot relative to the patient allows rapid volumetric data acquisition over large areas of the body. The data interpolation schemes developed for spiral CT permit the reconstruction of images with arbitrary position and spacing, including overlapping reconstruction for improved three-dimensional displays. Further, the improvement in rapidly covering larger volumes in spiral CT allows a better and more comfortable use of breath-holding techniques for improved imaging of internal organs that are susceptible to respiratory motion. Thus spiral CT, in particular, has improved the assessment of pulmonary nodules, has significantly increased the quality of three-dimensional surface-rendering type image representations, and—with the use of contrast agents—for the first time has allowed the display of vascular structures with previously unachievable quality.

³Dynamic CT is the acquisition of sequential images from the same anatomical area. Xenon CT measures the attenuation of certain tissue before and after the inhalation of xenon gas, which is dissolved in the blood stream, and thus allows one to quantitate brain perfusion.

The main limitation in spiral CT image quality today still is due to the available sustained power from the x-ray tube. While 40 kW are available typically for single scans, the power has to be reduced for spiral scans of 30 seconds and more to avoid overloading of the tube. As a result of the associated increase in quantum noise, soft tissue images frequently have to be reconstructed with smoothing convolution kernels. On the other hand, because of the high contrast of bone structures, the full spatial resolution can be made available for orthopedic examinations employing thin slice width settings. Blurring of the images in the direction of the patient's movement through the gantry can be minimized by appropriate choices for the table motion, and by reconstructing images for every 180° of scanner rotation rather than using the customary 360° scan coverage. Depending on the specific imaging situation, table feed values of one to two times the slice thickness per 360° of scanner rotation are being used today.

Even from the limited clinical experience with spiral CT available today, it is fairly obvious that this scanning mode will become the standard technique by the year 2000. Single-slice scanning will then be the exception. Required technical improvements will have to come in the available sustained x-ray power, as well as in the advances in computation time expected through increased processor performance. Another important impact of future technological developments will come from the use of multislice detectors or planar detector arrays extending over a larger solid angle, thus allowing better use of the available x-ray power and covering larger volumes within a given time. Such detectors are considered a most important future development. Depending on the technology employed, these detectors possibly will also allow improvement of the spatial resolution in the scanner axis direction.

3.4 Electron Beam Techniques

With conventional techniques, scan times cannot be shortened arbitrarily, because considerable masses have to be accelerated and moved, and because higher continuous x-ray power is difficult to achieve with the current rotating anode tubes. A technical alternative to overcome these problems and facilitate subsecond scanning is the use of electron beam techniques to replace conventional x-ray tubes. With electron beam systems the patient is surrounded by stationary anode rings inside a cone-shaped vacuum chamber. An electron beam is accelerated to about 130 or 140 keV and electromagnetically focused and steered onto the stationary anode rings, which can be cooled directly with an appropriate cooling agent. The electron beam can be rapidly scanned over the target rings, allowing scan times of 50 to 100 ms per slice. Using EKG triggering—synchronizing the data acquisition with the patient's heart beat by deriving a trigger from an electrocardiogram (EKG) signal—these short scan times allow the acquisition of unblurred images at a specific cardiac phase. Respiratory gating can be similarly used for longer studies, including the diagnosis of the great vessels. Switching between multiple target rings combined with conventional table feed permits fast and efficient volume acquisition.

The most promising and fastest-growing application of ultrafast electron beam based CT systems is coronary artery screening for early detection and quantitation of the calcification associated with arteriosclerosis. This procedure requires no contrast agents. Customarily, 20 to 40 3-mm-thick EKG-triggered slices are acquired at a rate of one at every heartbeat to cover a 6- to 12-cm volume through the coronary arteries. In chest imaging, electron beam CT allows covering the entire volume of the lungs within one breath-holding interval. This technique enables more accurate diagnosis of such conditions as pulmonary nodules, pulmonary embolism, and diffuse pulmonary emphysema than can be obtained from plain chest x-rays or slower CT techniques.

For the current generation of electron beam CT systems, image reconstruction still requires significantly more time than does image acquisition. Further research into faster filtered backprojection implementations or the increased use of Fourier domain-based reconstruction methods, along with an increase in fairly inexpensive processor speed, should help to close this obvious gap in the near future. Similarly, as with conventional CT scanners, the development of advanced detector array technologies will improve the spatial sampling of electron beam type systems. The spatial separation of the target rings and detectors along the scanner axis leads to “cone beam” errors that can show up as rib shadows. Volumetric imaging and reconstruction techniques are being developed to overcome this problem.

3.5 Data Handling and Display Techniques

Improvements in data acquisition speed and spatial resolution in both conventional and electron beam CT inevitably will lead to a vast increase in the amount of data collected and the number of images reconstructed. Data reduction techniques have to be developed to allow physicians to extract and display the information that is relevant out of the vast amount of three-dimensional volume data. A fairly straightforward method is to reduce a large number of transverse, tomographic slices into a few three-dimensional views of the structures and organ boundaries considered important. A critical prerequisite for generating three-dimensional views quickly and efficiently is the development of supervised semiautomatic segmentation techniques to routinely identify organ boundaries for the three-dimensional

representation. The resulting three-dimensional models are used increasingly in surgical planning and, even more directly, in CT-guided interventional techniques. Particularly for surgery that follows complex trauma, deals with congenital abnormalities, or reconstructs after tumor resection, accurate three-dimensional models of the anatomy are providing the detailed information required for accurate surgical planning as well as serving as a reference during the surgery itself. A particular application of three-dimensional volume techniques in CT that recently attracted a large amount of interest is the assessment of vascular structures with CT angiography. Three-dimensional volume images are acquired using a spiral CT scan synchronously triggered with contrast agent injection. For data reduction and display, two methods have been applied so far. A surface-rendering technique based on threshold type segmentation of the images is used predominantly for the display of complex vascular structures involving large vessels such as in the thorax. Alternatively, maximum intensity projection—a volume-rendering technique that generates projection images from the highest intensity pixels along the projection rays—of the image data allows a good visualization of smaller vascularity as well as vascular wall calcifications. Both of these projection techniques allow the retrospective generation of views at arbitrary angles, a capability that is not possible with conventional angiography. Applications of CT angiography established to date range from the diagnosis of renal artery stenosis to examinations of complex vascular abnormalities and aortic aneurysms.

3.6 Research Opportunities

To summarize, the following areas of scientific or technological development appear to have the greatest potential for enhancing the applications and clinical usefulness of x-ray CT:

- Increased instantaneous and, even more importantly, sustained x-ray power capabilities in both conventional x-ray tubes and electron beam systems;
- New and innovative technologies for the required high-intensity x-ray source, beyond those currently available;
- Two-dimensional detector arrays encompassing a larger solid angle and allowing improved spatial resolution along the scanner axis, including the associated high-throughput data acquisition electronics;
- Decrease in image reconstruction times, for example, through higher-performance low-cost processors and the increased use of Fourier domain based reconstruction algorithms;
- Mathematical means for utilizing the information contained in scattered photons;
- Effective means for the correction of patient motion-related artifacts;
- Better and easier-to-use three-dimensional data reduction and visualization techniques;
- Monochromatic x-ray generators, a major challenge; and
- Detectors that provide information on angles of incidence and high energy resolution, perhaps requiring new ideas from physicists and mathematical scientists.

3.7 Suggested Reading

1. Boyd, D., Couch, J., Napel, S., Peschmann, K., and Rand, R., Ultrafast cine-CT for cardiac imaging, *Am. J. Cardiac Imaging* **1** (1987), 175–185.
2. Brundage, B., Detrano, R., and Wong, N., Ultrafast computed tomography: Imaging of coronary calcium in atherosclerosis, *Am. J. Cardiac Imaging* **6** (1992), 340–345.
3. Kalender, W.A., Quo vadis CT? CT in the year 2000, *Electromedica* **61** (1993), 30–39.
4. Kalender, W.A., and Polacin, A., Physical performance characteristics of spiral CT scanning, *Med. Phys.* **18** (1991), 910–915.
5. Krestel, E. (ed.), *Imaging Systems for Medical Diagnosis*, Siemens Aktiengesellschaft, Berlin and Munich, 1990.
6. Polacin, A., Kalender, W.A., and Marchal, G., Evaluation of section sensitivity profiles and image noise in spiral CT, *Radiology* **185** (1992), 29–35.

7. Rubin, G., Dake, M., Napel, S., McDonnell, C., and Jeffrey, B., Three-dimensional spiral CT angiography of the abdomen: Initial clinical experience, *Radiology* **186** (1993), 147–152.
8. Schäfer, C., Prokop, M., Nischelsky, J., Reimer, P., Bonk, K., and Galanski, M., Vascular imaging with spiral CT, in *Advances in CT II*, R. Felix and M. Langer, eds., Springer-Verlag, Berlin, Heidelberg (1992), 109–115.
9. Schiemann, T., Bomans, M., Tiede, U., Höhne, K.H., Interactive 3D- segmentation, *Proc. SPIE* **1808** (1992), 376–383.
10. Stanford, W., and Rumberger, J.A. (eds.), *Ultrafast Computed Tomography in Cardiac Imaging: Principles and Practice*, Futura Publishing Co., Inc., Mount Kisco, N.Y., 1992.

Chapter 4

Magnetic Resonance Imaging

1. Introduction: Principles of Magnetic Resonance Imaging
2. Hardware
 - Magnet Systems: Current Status and Opportunities
 - Pulsed-Field MRI Systems
 - Radiofrequency Coils for MRI
 - Gradients for Dynamic MRI
3. Dynamic MR Image Reconstruction
4. Applications of Dynamic MRI
 - Blood Flow
 - Diffusion Imaging
 - Other Tissue Parameters
 - Functional Brain Imaging
 - Microscopic Imaging

This chapter highlights areas of interest related to magnetic resonance (MR) technology and its applications, particularly applications that involve dynamic magnetic resonance imaging (MRI). Among the hardware systems of an MRI device, the magnet, radio-frequency (RF), and gradient systems deserve particular R&D attention. For example, the development of suitable high-temperature superconductors (HTSs) and their integration into MRI magnets represents one potentially fruitful area of interaction between solid-state physicists, materials scientists, and the biomedical imaging research community. High-speed imaging pulse sequences, with particular focus on functional

Figure here Figure 4.1. Resolution, expressed as the reciprocal of imaging voxel volume, achieved in MR brain images by means of a volume head coil in scan times of less than 20 minutes and SNR sufficient to delineate anatomic structures.

imaging, are discussed in this chapter, as are algorithms for image reconstruction; both are promising fields of research. Significant attention is given also to two applications for which MRI has unique potential: blood flow imaging and quantification, and functional neuroimaging based on exploiting dynamic changes in the magnetic susceptibility. Finally, this chapter highlights recent technological advances toward real-time monitoring of interventional and therapeutic procedures.

MRI technology has undergone amazing strides over the last two decades, much of it due to advances from the mathematical sciences and physics. For example, Figure 4.1 demonstrates a tremendous improvement in resolution over that period. There is every reason to believe exciting progress still lies ahead.

Figure here Figure 4.2. Time domain nuclear magnetic resonance signal from volume element $dx dy dz$ in an object of magnetization density $M_{xy}(t)$ in the presence of a spatial encoding gradient \mathbf{G} .

4.1 Principles of Magnetic Resonance Imaging

Unlike its x-ray counterparts, magnetic resonance imaging (also known as nuclear magnetic resonance (NMR) imaging) is not a transmission technique. Rather, the material imaged is itself the signal source (i.e., the macroscopic spin magnetization \mathbf{M} from polarized water protons or other nuclei, such as ^{23}Na or ^{31}P). The motion of the magnetization vector of uncoupled spins, such as those for protons in water, is conveniently described in terms of the phenomenological Bloch equations:

$$\frac{d\mathbf{M}}{dt} = \gamma\mathbf{M} \times \mathbf{H} - \frac{M_{xy}}{T_2} - \frac{M_z - M_0}{T_1}, \quad (4.1)$$

where γ is the gyromagnetic ratio, \mathbf{H} the effective field, M_0 the equilibrium magnetization, and T_1 and T_2 the relaxation times. T_1 is the characteristic relaxation time for longitudinal magnetization to align with the magnetic field: following a perturbation such as a 90° RF pulse, the longitudinal magnetization typically returns to its equilibrium value, M_0 , with a time constant T_1 . Likewise, T_2 is the characteristic time for decay of coherent magnetization in the transverse plane: the transverse magnetization decays exponentially with time constant T_2 to its equilibrium value, $M_{xy}^0 = 0$. Both relaxation times are determined by the interaction of water or other nuclei with macromolecules in tissues. T_1 and T_2 contribute independently to the contrast between different tissues.

There is, in general, no closed-form solution to equation 4.1 (although section 14.1.6 introduces two approximate solutions). Ignoring the relaxation terms (which can be done without loss of generality when formulating the imaging equations), the steady-state solution of equation 4.1 in the presence of a static polarizing field $H_0 = H_z$ corresponds to a precession of the magnetization about the field at a rate $\omega_0 = -\gamma H_0$. Since the detected signal voltage is proportional to $d\mathbf{M}/dt$ and T_1 is on the order of 1 s, only components transverse to the polarizing field give rise to a detectable signal.

At its core, MRI exploits the field dependence of the precession frequency by superimposing a magnetic field gradient $\mathbf{G} = (\partial H_0/\partial x, \partial H_0/\partial y, \partial H_0/\partial z)$ onto the static polarizing field $H_0 = H_z$ to spatially encode information into the signal. In this manner, the resonance frequency ω becomes a function of spatial position \mathbf{r} , according to

$$\omega(\mathbf{r}) = -\gamma(H_0 + \mathbf{G} \cdot \mathbf{r}). \quad (4.2)$$

If precession due to the polarizing field (the first term in equation 4.2) is ignored, the complex MR signal is seen to evolve as $\exp(-i\gamma\mathbf{G} \cdot \mathbf{r}t)$. Following excitation by an RF pulse, and ignoring relaxation, the time-dependent signal $dS(t)$ in a volume element $dx dy dz$ becomes (Fig. 4.2)

$$dS(t) = M_{xy}(\mathbf{r}) \exp[-i\gamma(G_x x + G_y y + G_z z)t] dx dy dz. \quad (4.3)$$

It is convenient to express the signal as a function of the spatial frequency vector

$$\mathbf{k}(t) = \gamma \int_0^t \mathbf{G}(t') dt', \quad (4.4)$$

where $\mathbf{G}(t')$ is the time-dependent spatial encoding gradient defined above, which imparts differing phase shifts to spins at different spatial locations. Equation 4.4 shows that the time variation of the spatial frequency vector is determined by the integral over the gradients; the time sequence of the RF pulses and magnetic field gradients is known as the “pulse sequence” of the MRI acquisition.

Rather than expressing the MRI signal in terms of the magnetization, which is modulated by the relaxation terms in a manner specific to the excitation scheme used, it is customary to describe the signal as a function of the density of spins in the tissue. For an object of spin density $\rho(\mathbf{r})$ the spatial frequency signal $S(\mathbf{k})$ thus is given by

$$S(\mathbf{k}) = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \rho(\mathbf{r}) e^{-i\mathbf{k}(t) \cdot \mathbf{r}} d\mathbf{r}, \quad (4.5)$$

with integration running across the entire object. Pictorially, the spatial frequency may be regarded as the phase rotation per unit length of the object the magnetization experiences after being exposed to a gradient $\mathbf{G}(t')$ for some period t .

Figure here Figure 4.3. Simplified block diagram of a typical MRI system.

One further recognizes from equation 4.5 that $S(\mathbf{k})$ and $\rho(\mathbf{r})$ are Fourier transform pairs, and thus

$$\rho(\mathbf{r}) = \frac{1}{2\pi} \int_{\mathbf{k}_1}^{\mathbf{k}_2} S(\mathbf{k}) e^{i\mathbf{k}(t) \cdot \mathbf{r}} d\mathbf{k} . \quad (4.6)$$

Following an RF pulse, the transverse magnetization is subjected to the spatial encoding gradient $\mathbf{G}(t)$, which determines the spatial frequency vector $\mathbf{k}(t)$ according to equation 4.4. The manner in which k-space is scanned and the path of the k-space trajectory are determined by the waveform of the gradients $\mathbf{G}(t)$ and the sequence of RF pulses.

In contrast with the signal used in computed tomography, the MRI signal is sampled in the spatial frequency domain, rather than in object space. The most common technique involves alternate application of two or three orthogonal gradients. Signal detection and sampling in the presence of a constant, so-called readout gradient applied along one dimension then produces a rectilinear grid of data from which the image pixel amplitudes are obtained by the discrete Fourier transform. Other imaging schemes involve radial or spiral coverage of k-space.

A block diagram of a typical MRI apparatus is given in Figure 4.3. The heart of the apparatus is the magnet system, typically a superconducting solenoid. The second system is the transmit/receive assembly, typically consisting of a transmitter and power amplifier to generate the RF voltage to be fed into the transmit RF coil, thereby creating a circularly polarized RF field. The latter produces transverse magnetization, which in turn induces an RF voltage in the receive coil (which may actually serve for both transmission and reception). The ensuing signal is amplified and demodulated in the receiver electronics. Unique to the MRI device is the gradient system, which permits generation of the previously defined time-dependent gradient fields needed for spatial encoding. Both the transmit/receive and gradient systems are under the control of a data acquisition processor that ties into the main central processing unit.

4.2 Hardware

4.2.1 Magnet Systems: Current Status and Opportunities

The core of an MRI apparatus is the magnet that generates the field for nuclear polarization. The magnetic field for MRI must be extremely stable (0.1 ppm per hour) and uniform (10 to 50 ppm in a sample volume of 30- to 50-cm diameter). Magnet types in current use are of the superconducting, resistive, and permanent design. The large majority of MRI units use superconducting magnets, which provide fields of high strength and stability. They are, however, more expensive to manufacture and are cost-effective only for fields of 0.3 tesla (T) and higher. Most currently produced magnets are based on niobium-titanium (NbTi) alloys, which are remarkably reliable. These typically require a liquid helium cryogenic system with regular replenishment of the cryogens to keep the conductors at superconducting temperature (approximately 4.2 K).

In recent years the cryogenic efficiency of such magnets has been improved so that many systems now require replenishment at 9-month intervals only, compared to monthly intervals 10 years ago, and approaches eliminating the liquid helium and using conduction cooling have been proposed. Liquid nitrogen, which previously was replenished weekly, has now been completely eliminated in some systems in favor of electrically driven refrigerators. Advances in active shielding of external magnetic fields, together with creative magnetic, mechanical, and cryogenic designs, have reduced the physical volume of these magnets by typically 40% over the past decade, and the once-delicate cryostats are now rugged enough so that mobile MRI installations can travel the world's highways without incident.

Permanent and resistive magnets, both iron core and air core, are cost-effective for field strengths below approximately 0.5–0.3 T. Permanent magnets, in particular, can have very low operating costs. While permanent magnets have a weight penalty, novel designs have reduced this as well, particularly when compared to the total weight of the magnet and passive ferromagnetic shield in some other designs.

Because of its ductility and ease of fabrication, the solid-state alloy of NbTi is the usual conductor used in superconductivity applications. The only other alloy used commercially is Nb₃Sn, a brittle compound but one whose higher critical temperature allows it to be used in higher-field systems or at temperatures higher than 4 K. It has been used in several novel conduction-cooled open-architecture MRI magnets developed to test interventional and therapeutic applications of MRI (see Chapter 12). These systems operate at approximately 10 K. A natural next step in the evolution of MRI technology might be to wind MRI magnets with a high-temperature superconductor (HTS), one of a class of materials with critical temperatures above approximately 25 K. The first of these materials were

developed by Bednorz and Muller in 1986, and ones with critical temperatures as high as 164 K have been reported. Basically ceramic alloys, these materials are constructed with powder metallurgical or thin-film techniques and are inherently brittle (similar to Nb_3Sn). Considerable strides have been made in the manufacture of long production lengths of bismuth-strontium-calcium-copper oxide/silver (BSCCO/Ag), which can carry higher currents than can the low-temperature superconductors at 4.2 K. In general, HTSs appear capable of carrying more current than do low-temperature superconductors at the latter's operating temperatures in fields over 14 T.

A critical component of the superconducting magnet is the leads that are required to initially energize the magnet. If the magnets are persistent, these leads are retracted in order to reduce to a minimum the heat load on the system. If the magnets are conduction cooled, the power leads and any instrumentation connected to the coil are situated in the vacuum space with the windings. Because of the difficulty of disconnecting such leads in a vacuum, conduction-cooled coils usually use permanently connected leads, creating a significant heat load. There is promise that using HTS electrical leads could reduce such loads because of their natural low thermal conductivity.

Another critical technology is the manufacture of true zero-resistance joints for connecting individual lengths of the magnet coils and for the persistent switch used to disconnect the magnet coils once the magnet has been energized. Creation of such components for NbTi magnets is routine but is more involved for Nb_3Sn , although the use of this material in NMR spectrometers indicates that the joint technology does exist.

4.2.2 Pulsed-field MRI Systems

MRI systems require a relatively high magnetic field for polarizing the magnetic moments to provide an adequate signal for signal-to-noise ratio (SNR) considerations. This polarizing operation, of itself, is relatively noncritical. Large variations, of the order of 10%, can readily be tolerated since such a variation merely shades the resultant image and can be easily compensated for if desired. However, when this same field is used for the readout operation, as is done in all existing systems, the requirements are drastically different. The field must now have homogeneities of the order of 1.0 ppm to avoid signal loss and distortion. The signal loss stems from variations in frequency within an imaging voxel¹ where the resultant destructive phase interference causes enhanced signal decay, significantly reducing the sampling time available. These problems are particularly severe in high-speed systems that attempt to cover large amounts of k-space following each excitation. The limited readout time restricts the amount of coverage and the SNR, making real-time performance impractical.

In pulsed-field systems the functions of polarization and readout are separated, enabling a number of significant performance and cost advantages. A pulsed field is applied and left on for a time comparable to the relaxation time T_1 , to polarize the magnetic spins. Since the magnet has considerable thermal inertia and is on only during the imaging interval, the peak field can be made relatively high, of the order of 1.0 T, without excessive temperature rise. Thus the very use of a pulsed field has a significant SNR advantage over static resistive magnets. In addition, the time of the pulse can be used for providing T_1 contrast by determining the degree of polarization buildup for the different materials. Of course T_2 contrast is established in the usual way by varying the time over which the origin of k-space is scanned during the readout. As indicated, this pulsing operation is highly noncritical as to homogeneity. It therefore enables the use of low-cost simple magnets. The magnets now being used are either very large solenoids or vertical field magnets with pole pieces much larger than the imaged region. These have high stored energy and result in very limited accessibility to the patient for guided biopsies and/or therapeutic procedures. Pulsed magnets enable the use of simple configurations such as a pair of coils or a "C"-shaped core having a cross section comparable to the region of interest. This simplicity enables the construction of specialized machines for joints, head, breast, heart, and so on. The readout bias field following the pulse is made very low so that its variations will result in negligible T_2^* or distortion considerations.² Susceptibility-induced variations (which scale with readout field strength) should be negligible.

Although the readout bias field is relatively noncritical, it can be a major consideration in a simple low-cost design. For example, if the readout field were 1,000 times smaller than the polarizing field, for comparable performance its homogeneity should be 1 part per 1,000 as compared to 1 ppm. Although this ratio is readily realized in a large solenoid, it remains quite difficult to achieve in a simple configuration such as a pair of coils positioned on opposite sides of the body. To achieve essentially complete immunity to inhomogeneity of the readout bias field, comparable to the immunity of the pulsed polarizing field, an oscillating bias field is used.

Perhaps the biggest problem created by this approach is the radio-frequency receiving system. The desired result is to obtain an SNR commensurate with the size of the polarizing field. This will happen only if the noise generated is due primarily to body losses, where other losses, including those of the coil, are negligible. The problem is that,

¹ Volume element, the higher-dimensional analog of the more-familiar pixel (picture element) that is the basic unit of a planar image.

² T_2^* is the effective transverse relaxation time, which determines the rate of decay of the signal in the presence of magnetic field inhomogeneity.

at very low readout frequencies, the body losses are significantly reduced, thus making coil noise relatively more important. As a result systems using very low readout frequencies require low-loss coils. These can include coils made of cooled materials and high-temperature superconductive coils. The use of these low-loss, high-quality-factor coils introduces another problem, that of insufficient bandwidth.

The small signal also results in an increased shielding problem since external interfering signals are more significant. One solution to this problem is the use of gradiometers, coil configurations that respond solely to field variations and ignore uniform fields arising from more remote sources.

4.2.3 Radio-frequency Coils for MRI

Radio-frequency coils constitute the essential means of stimulating the spin systems and are the doorway for the sensing of MR data. Several research opportunities are presented by the need to improve the quality of the data and the rate of data acquisition. The data quality could be improved with better coil designs and, in some applications, cooled or superconducting coils, which result in lower noise. Better data quality could also be facilitated with multiple-acquisition coils operating in parallel.

Computational Design of RF Coils Historically, RF coils have been constructed according to basic design principles learned through experience by the individual investigator. More sophisticated computational RF coil design is an area for future growth, for several reasons:

- MRI is an intermediate field problem, and such fields are difficult to simulate even for quasi-static analyses.
- Very high frequencies are beyond the quasi-static regime of Maxwell's equations. Wavelength effects inside the patient are important and must be considered. Designing transmitters and receivers based on an understanding of dielectric and wavelength effects, so as to enable uniform excitation and detection, is beyond the scope of quasi-static arguments.
- Accurate estimates of power deposition in patients are needed for high-field, high-power imaging sequences. Since neither experimental measurements nor first-order calculations can accurately predict local power deposition, accurate field calculations are required as a basis for developing safe operating guidelines.
- Very high frequency coils are still tuned and matched by trial and error. As in the design of lower-frequency resonators, it would be extremely useful to have a set of independent controls for the coil's different oscillatory modes that produce circularly-polarized fields.

All of these applications require the availability of a true three-dimensional solution to the time-varying fields. As yet, there has been little published on MRI applications of finite-element calculations. A few abstracts have described some small-scale problems, but none have attacked a full size problem. Generally this work has not taken a full three-dimensional approach; usually one dimension in the problem has been very coarsely sampled in the analysis.

Assuming that local heating effects in a body need to be quantified, a finite-element calculation with elements on the order of 1 cm^3 would be needed. A head-size coil in a shield might require something on the order of 100,000 to 200,000 elements. The scale of this problem is appropriate for even a modest modern workstation; with an iterative finite-element solver, something on the order of a week of CPU time might be required. This scale is just right for a research group to begin investigating. National supercomputer facilities exist that could greatly reduce this time.

Beyond the physics of solving electromagnetic problems, issues specific to MRI complicate the interpretation of results. The best methods for visualizing three-dimensional field values are not obvious. Development of such techniques will require computational tools specific to MRI, for example, means of performing a simulated MR pulse sequence and visualizing the images created by it.

Cooled Receiver Coils for MR Imaging Most SNR calculations in MRI rest on the basic assumption that the patient contributes the dominant part of the noise. This assumption is known to be incorrect in two important cases:

- When coil loading increases rapidly with increasing frequency. Consequently in low-field MRI, patient contributions to noise may not dominate.
- When coil loading decreases rapidly with the length scale of the system. As the receiver coil is made smaller, the contribution from patient loading can be made smaller than coil contributions at any field strength. For microscopy applications the coil is the dominant source of resistance.

High-temperature superconducting materials that contribute little noise when operated at liquid nitrogen temperatures deserve to be evaluated for use in MRI receiver coils. The most promising technology involves thin film HTSs. These materials can be used in large static fields provided that the plane of the thin film is parallel to the field. The limitations of thin film technology are those imposed by the vapor deposition process used to create them. Very large coils are at present not possible; the practical upper limit is about 8 cm. Alternative possibilities are tape HTSs or the use of conventional copper coils at liquid nitrogen temperature.

Use of Multiple Receivers While the use of multiple acquisition coils operating in parallel could improve data quality, the technological requirements for a practical multi-coil array are formidable: beyond the issues of building a decoupled array of surface coils, there are considerable demands on the system hardware to be able to deal with data rates that are an order of magnitude higher than those from a conventional receiver coil. A flexible and powerful system architecture is needed that allows for the acquisition and manipulation of larger data streams. An active research area is the multiplexing of several RF channels through one wideband receiver, digitizer, and computer. Multiplexing is possible in both the frequency and time domains. Noise crosstalk between channels must be prevented by tight filtering near the RF coils. Crystal filters and filters built with HTS high current resonators are contenders for this task.

The simplest approach for the combination of data from the array is based on a linear combination of the individual images into a composite image. Provided the receiver coil sensitivities are known and a collection of equal-noise images exists, then the least-noise combined image is a sum of images weighted by their sensitivities. However, achieving reliable and noise-free estimators of individual coil sensitivities remains a challenge that calls for more sophisticated techniques, possibly statistical ones.

Alternatively, there are several adaptive techniques for image combination that act strictly on the original images without any need for filtering. Principal component analysis (PCA) can take a collection of original images with different noise levels and synthesize a composite image that is the noise-optimum linear combination. A collection of surface coil images requires a linear combination in which the linear weighting is spatially dependent. To calculate something like a PCA combination, one would require a means of performing an eigenvector analysis on a small segment of the entire image, repeating this process across the image, and then interpolating between the different subimages. Using the interpolation scheme is essentially the same as working from estimators based on filtered images, but the advantage is that an assumption of equal noise levels is no longer necessary. One could then produce an optimal linear combination of surface coil images derived from input images with greatly different noise characteristics.

Multiple coil technology might also be used to speed up the MR imaging sequence. One can argue from simple heuristic principles that acquisition of signal from two (or more) receiver coils yields twice as much information per phase encoded signal and that we ought to be able to use the additional information to reconstruct an image based on half the number of phase encoded signals. One method of implementing this uses receiver coils constructed with a known sensitivity. It requires that one coil be a standard uniform receiver of the type common in MRI; the second is an RF gradient coil. Fortunately these coil sensitivities are easy to implement in a conventional horizontal-field superconducting magnet: one coil is the standard Helmholtz pair, and the second is a Helmholtz pair with opposed windings. An imaging acquisition consists of a standard phase encoded sequence, but every other acquisition is skipped. The skipped phase encoded echoes can be calculated from the acquired data on nearby steps in Fourier space from each of the separately received data sets. Possibilities for extending these ideas to multiple receiver coils provide interesting research challenges.

4.2.4 Magnetic Field Gradients

Spatial encoding in MRI typically uses spatial variations in the static or radio-frequency magnetic field. Because such variations are small compared to those in the static magnetic field B_0 used for MRI, a set of “gradient coils” can be used to produce a very nearly linear gradient along any direction. Gradients in the radio-frequency magnetic field, although useful for some purposes, are not discussed here.

Besides being the essential element for spatial encoding, the gradient system of the MRI scanner is responsible for encoding dynamic information related to flow, diffusion, or modulation of magnetization for spatial tagging. In many applications, the quality of the data obtained is currently limited by hardware-imposed limitations on the gradient amplitude and/or rate of change (slew rate). As an example, such limitations hamper echo-planar imaging (EPI), where ideally all the k-space data would be acquired following a single excitation over a period on the order of T_2 . Limited gradient amplitude and switching rates currently restrict the temporal and spatial resolution of EPI.³

³For background on EPI see, for example, the paper by Cohen and Weisskoff listed in the suggested reading for this section.

A second example is diffusion imaging, which is limited primarily by gradient amplitude to very long echo times on conventional systems. When high-performance gradient hardware is available, some applications are significantly limited by restrictions on maximum permissible field slew rates in humans ($dB/dt < 20$ T/s). Such restrictions are in place to prevent nerve stimulation by induced eddy currents.

Local versus Whole-Body Gradients The MRI scanner's gradient system consists of two major components, the gradient coil, which is a set of windings that determines the spatial distribution of current, and the gradient power supply, which typically is a digitally controlled audio amplifier that supplies power to the coil. Stronger and faster switching gradients can be achieved by improving either of these components. Local gradient coils (LGCs), which are smaller than the whole-body gradient coils that are built into clinical imaging systems, are designed for specific purposes or areas of anatomy. Their smaller size gives them three advantages over whole-body gradients. The first is higher performance in the form of stronger gradient amplitudes per unit current, and higher gradient slew rates per unit voltage. The second is that because of their smaller size, LGCs can create higher gradient slew rates for a given imposed limit on field slew rate—a more fundamental advantage in that LGCs, in the absence of power limitations, can achieve stronger gradients with faster switching than can whole-body gradient systems. The third advantage is lower cost, now approximately an order of magnitude lower.

High-powered whole-body gradient systems generally fall into one of three categories: high-powered linear amplifiers, switching systems, and resonant systems. High-powered linear systems are self-explanatory. Both switching systems and resonant systems use energy storage techniques to minimize power dissipation. Switching systems generally use semiconductor switches to step between current states very rapidly, while resonant systems use an L-C circuit with the gradient coil as the inductor to efficiently generate sinusoidal current waveforms. A few of these approaches have been used successfully to design and implement fast imaging systems capable of, for example, single-shot echo-planar imaging. The primary advantage of this approach is that the hardware is not specific to the anatomic region of interest, allowing imaging of various regions without requiring a change of gradient hardware.

The remainder of this section focuses on LGCs.

Design Considerations Many factors must be taken into account in the design of LGCs, including coil geometry, coil size, anatomy of interest, desired gradient strength, efficiency (gradient per unit current), inductance, desired slew rate, eddy currents, gradient uniformity, forces and torques, heat dissipation, and nerve stimulation.

Geometry: The geometry of LGCs, like that of local RF coils, can be divided into three categories: whole-volume coils, which completely surround the anatomic structure of interest; partial volume coils, which partially surround the structure; and surface gradient coils, which are placed against one side of the structure.

The advantages gained by the use of an LGC depend strongly on the size of the coil. For a coil of a given design, with a fixed number of turns, the efficiency scales inversely with the size squared while the inductance scales with the size, indicating that the size of the coil should be as small as possible. The lower limit on the size of the coil is set by three other factors: the coil must have a region of usable gradient field uniformity that encompasses the anatomic region of interest; the coil itself must be large enough to accommodate the anatomic region of interest; and sufficient room must be left for the RF structure and return flux to allow dominant loading of the RF coil by the subject. The largest of these three lower limits determines the optimum size of the coil.

Eddy Currents: Eddy currents are currents that are induced in conducting structures due to time-varying magnetic fields. Gradient eddy currents effectively limit the field slew rate by creating transient fields that usually oppose the desired gradient fields. Structures in an MRI scanner that may support eddy currents include the cryostat, the RF coil, and the RF shield. For whole-body gradient coils, eddy currents generated in the cryostat are usually of the greatest concern because the cryostat is physically close to the gradient coil and has very low resistance. In general, LGCs are much smaller and therefore couple much less strongly to the cryostat. However, RF coils or shields that have low-resistance closed-DC current paths in close proximity to the gradient coil can support eddy currents and degrade the gradient field waveforms. This problem must be considered carefully in the design of RF coils and shields for use with LGCs.

Linearity: Because the gradient fields directly map the spatial coordinates of physical space to the coordinates of the image, perfect gradient field uniformity is desired. However, if the gradient fields are not perfectly linear but are known, spatial transformations can be applied to remove warp from the images. This process can partially alleviate the problems associated with using nonlinear gradients, but several problems remain. In two-dimensional imaging, the most common type, if the slice-select gradient is nonlinear, then information is obtained from tissue outside the desired slice and the slice cannot be correctly reconstructed. If three-dimensional data are collected, or if the slices are contiguous, then three-dimensional spatial transformations can be used. These transformations are straightforward but increase image reconstruction time. Because nonlinearities in the gradients cause variations in

the voxel sizes, the signal intensity, SNR, and spatial resolution will be modulated by these variations, potentially adding to the difficulty of image interpretation.

Lorentz Forces: The currents carried by a gradient coil interact with the static field, causing the coil to experience large forces. In a uniform static field, a closed current loop cannot experience net forces, but it can experience net torques. If these torques are not internally balanced, the coil must be mounted in a very strong and rigid mechanical structure that is tied to the structure of the main magnet in order to prevent large-amplitude vibrations of the coil. Torques on an unbalanced head coil in a 1.5-T field are on the order of 100 N-m. Internal balancing of the torque in a coil can be achieved by symmetry or by explicit nulling of the torque moment in an asymmetrical design.

Design Techniques: Numerous mathematical techniques have been used to design gradient coils, many of which are described in a 1993 review article by Turner listed in the recommended reading below. Most designs to date describe current patterns on the surface of one or two cylinders, as this is the most natural geometry for whole-body patient access, and for LGCs a natural geometry for the head and extremities. Early designs consisted of simple pairs of loops for longitudinal gradients and saddle coils for transverse gradients. Now, a more elegant approach has been introduced that explicitly minimizes inductance while creating the desired fields within the region of interest. Inductance is a quantity that is difficult to minimize using other techniques because it is a nonlinear function of the current density. For a given coil geometry, an appropriate transform applied to the current density linearizes the inductance, greatly simplifying the problem of inductance minimization. For planar and cylindrical geometries, simple two-dimensional Fourier transformation serves this purpose. In the transformed space the inductance is a linear function of the transform components, and a Lagrange multiplier technique can be used to satisfy the desired field constraints while minimizing inductance. After deriving a current density pattern in transform space, a reverse transform is applied to produce a current density pattern in physical space, and approximation by discrete current paths gives a final design. Several variations of the minimum inductance technique have been introduced to apply the technique to different geometries and to use the same principles to minimize power dissipation instead of inductance.

Another class of design techniques uses numerical methods to optimize combinations of gradient uniformity, inductance, and coil efficiency. Using either a parameterized or a point-by-point description of the current path, mathematical techniques such as gradient descent, simulated annealing, Monte Carlo, and simplex can be used to minimize a chosen cost function. Numerical methods are in general more computationally intensive than analytical techniques, but they allow for incorporation of arbitrary constraints in a simple manner, and they can directly produce final designs, avoiding errors that may be introduced by the approximation of continuous current densities by discrete wires.

Applications Besides being an integral part of the spatial encoding process, high-performance gradient systems have many other applications. In angiography, stronger and faster gradients allow for shorter flow-compensating gradients and hence shorter echo times, T_E . These conditions decrease flow-related dephasing and increase the detected signal. In saturation tissue tagging, such as myocardial tagging to observe cardiac wall motion, stronger gradients can improve tag profiles and decrease tagging time, thus generating sharper tagged grids in the final images. For imaging of very short echo time T_2^* regions, it is necessary to employ strong gradients in order to ensure that the imaging gradients dominate the static field inhomogeneities. Short T_2^* occurs primarily in areas with strong susceptibility gradients such as in lung tissue.

Bioeffects For MRI in humans, the fundamental limitation on the performance of gradient systems is the need to avoid nerve stimulation. For the application of single-shot echo-planar imaging (EPI) of the head, for example, a three-axis gradient coil enables the use of EPI in an otherwise conventional commercial scanner. Using standard gradient amplifiers, the 20-T/s limit imposed by the U.S. Food and Drug Administration can be exceeded with this coil, so that EPI is no longer hardware limited. A part of the coil optimization question that has not been addressed specifically is how to design a coil with appropriate field distribution, access, efficiency, and so on that also explicitly minimizes the peak field to which the subject is exposed. This is an added dimension to the coil design problem that will have a different solution for every imaging region of interest and every anatomical site.

4.2.5 Research Opportunities for MRI Hardware

Magnet Systems

- Development of *economical* higher-temperature superconducting magnets, using Nb_3Sn and other higher-temperature materials.
- Development of designs for *economical* magnets for special applications, including specific anatomic parts, or for special disciplines such as therapy.

- Development of designs for more economical magnetic field shielding that would allow retrofitting of magnets into existing diagnostic, interventional, and surgical rooms.

Pulsed-field MRI

- Development and validation of strategies for signal readout that minimizes introduction of additional noise and interference.
- Development of means for energy recovery during collapse of the polarizing field.

RF Coils

- Design of uniform transmitters and receivers that include dielectric and wavelength effects.
- Development of methods for three-dimensional modeling of RF fields.
- Design and evaluation of high-temperature superconducting coils and associated cooled pre-amplifiers for low-field imaging.
- Design of SNR-efficient high-speed image combination and reconstruction techniques for multicoil arrays.

Gradient Systems

- Design of very short head gradient coils using current return paths at a greater diameter than that of the primary windings.
- Design of acoustically quiet gradient coils.
- Design of head gradient coils with good subject access for visual and auditory task presentation.
- Optimization of head immobilization devices compatible with head gradient coils.
- Development of methods for electrical decoupling of RF coils and gradient coils in close proximity.

4.2.6 Suggested Reading Related to MRI Hardware

Magnet Systems

1. Stekly, Z.J.J., and Gregory, E., in *High-Temperature Superconducting Materials Science and Engineering*, Pergamon Press, Tarrytown, N.Y., 1994.
2. Stekly, Z.J.J., and Gregory, E., in *Intermetallic Compounds: Principles and Practice*, J. Westbrook and R.L. Fleisher, eds., John Wiley & Sons, New York, 1994.

Pulsed-field MRI

3. Macovski, A., and Connoly, S., Novel approaches to low-cost MRI, *Magn. Reson. Med.* **30** (1993), 221–230.
4. Mansfield, P., and Morris, P., *NMR Imaging in Biomedicine*, Academic Press, New York, 1982.
5. Morris, P.G., *Nuclear Magnetic Resonance Imaging in Medicine and Biology*, Oxford University Press, New York, 1986.

RF Coils

6. Carlson, J.W., and Minemura, T., Imaging time reduction through multiple coil data acquisition and image reconstruction, *Magn. Reson. Med.* **29** (1993), 681–688.
7. Ra, J.B., and Rim, C.Y., Fast imaging using subencoding data sets from multiple detectors, *Magn. Reson. Med.* **30** (1993), 142–145.
8. Roemer, P.B., Edelstein, W.A., Hayes, C.E., Souza, S.P., and Mueller, O.M., The NMR phased array, *Magn. Reson. Med.* **16** (1990), 192–225.
9. van Heteren, J.G., James, T.W., and Bourne, L.C., Thin film high temperature superconducting RF coils for low field MRI, *Magn. Reson. Med.* **32** (1994), 396–400.

Gradient Systems

10. Cohen, M.S., and Weisskoff, R.M., Ultra-fast imaging, *Magn. Reson. Imaging* **9** (1991), 1–37.
11. Kwong, K.K., Belliveau, J.W., Chesler, D.A., et al., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation, *Proc. Natl. Acad. Sci. USA* **89** (1992), 5675–5679.
12. Ogawa, S., Tank, D.W., Menon, R., et al., Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging, *Proc. Natl. Acad. Sci. USA* **89** (1992), 5951–5955.
13. Rosen, B.R., Belliveau, J.W., and Chien, D., Perfusion imaging by nuclear magnetic resonance, *Magn. Reson. Quart.* **5** (1989), 263–281.
14. Rzedzian, R.R., and Pykett, I.L., Instant images of the human heart using a new, whole-body MR imaging system, *Amer. J. Roentgenol.* **149** (1987), 245–250.
15. Turner, R., Gradient coil design: A review of methods, *Magn. Reson. Imaging* **11** (1993), 903–920.
16. Turner, R., Minimum inductance coils, *J. Phys. E: Scientif. Instrum.* **21** (1988), 948–952.
17. Turner, R., Optical reconstruction of NMR images, *J. Phys. E: Scientif. Instrum.* **18** (1985), 875–878.
18. Turner, R., Jezzard, P., Wen, H., et al., Functional mapping of the human visual cortex at 4 and 1.5 tesla using deoxygenation contrast EPI, *Magn. Reson. Med.* **29** (1993), 277–279.

4.3 Dynamic MR Image Reconstruction

Dynamic magnetic resonance imaging today can be interpreted to mean either very fast or essentially continuous scanning. In practice, dynamic means fast enough to observe an object before its properties change: fast enough to image the heart before it moves or fast enough to image a tissue that is taking up a contrast agent before its relaxation properties change. As such, dynamic imaging requires the fastest imaging schemes possible. So, how does image reconstruction enter the picture? From the simplest perspective, the fastest possible reconstruction scheme is desired, which today is probably the fast Fourier transform (FFT). However, when data are collected quickly in an MRI experiment, the signal-to-noise ratio (SNR) is reduced. So a reconstruction method with the highest possible SNR is desired as well. Another factor is the resolution. Physical characteristics will also affect the choice of resolution. Finally, and most importantly, how do all these needs affect the way the MRI data themselves are collected so as to optimize the image quality and information content? Can the concepts of resolution, SNR, and speed be uncoupled? Clearly, this image reconstruction problem represents a challenge involving both the mathematics and the physics of MRI.

In this brief outline of research opportunities related to real-time MR image reconstruction, several approaches are considered: partial Fourier reconstruction, parametric estimation, Bayesian analysis, wavelets, image filtering, echo-planar imaging, and the general question of k-space coverage. One particularly important point to make here is that most of these methods have focused on one-dimensional problems with some brief forays into higher dimensions. In MRI these higher dimensions provide a wealth of data that require alternative methods of reconstruction.

4.3.1 Partial Fourier Reconstruction

In the simplest version of MR image reconstruction, data are collected in a complex form symmetrically about the origin. In k-space notation, the data run from $-k_{\max}$ to k_{\max} in discrete steps. For example, when the FFT is used, $2n$ points are collected for $-n \leq i \leq n-1$ with $k(i) = i\Delta k$. To save time, data can be collected in principle for just $0 \leq i \leq n-1$ and, when the object to be reconstructed is real, the data can be complex-conjugated and a conventional complex FFT done to recover the real object. In practice, the reconstructed image data are complex-valued due to factors such as main field inhomogeneity and flow effects. Discussed below are four specific strategies that attempt to deal with this more general situation. These discussions ignore Gibbs ringing, but it is included in section 4.3.2.

Predominantly One-sided Data Collection The most common way used to save time is to collect just m points before the echo and n points after the echo. The time to acquire the data is thus reduced to a fraction $(m+n)/2n$. The loss in SNR is the square root of this fraction. Unfortunately, when the object is complex, the operation of complex conjugation is foiled. One approach used was to assume that the object had only low-spatial-frequency phase variations that were well defined within the data points $-m \leq i \leq m$. A low-spatial-frequency image was then reconstructed and the phase from this image used to rotate the additional information into the real part by

the above convolution. This common procedure is technically correct only for small phase angles. In practice, a projection onto convex sets (POCS) approach can be used to obtain a more stable solution to $\rho(x)$.

The procedure can actually be restated as a matrix problem and a solution to the inverse can be found. The disadvantage to this approach is that matrix inversion is time-consuming. The advantage is that the ill-posedness of the problem can be better understood. To date, though, the general problem of correcting for a high-spatial-frequency phase error has not been fully explored. The reconstruction errors that manifest themselves in the presence of these phase errors show up as an oscillatory signal response about the region where the error has not been accounted for. These reconstruction errors can be quite large and often cause bright or dark spots in the MR image. They are thus very detrimental to the use of one-sided data collection, making quite problematic its application to data obtained with techniques that tend to produce phase errors, such as gradient echo sequences or sequences that are not flow compensated.

Predominantly Every Other Point There are indeed many other ways to attempt to optimize the collection of the data to deal with the above problem, but one that is of particular interest is that which collects the same data in the center (i.e., from $-m \leq i \leq m$) and then from all odd points for $-n \leq i < -m$ and all even points for $m < 1 \leq n-1$. Collecting the data in this way has the advantage that high-spatial-frequency information is collected throughout what would have been the conventional data set and hence can accommodate field inhomogeneity errors that are either positive or negative. The blurring artifacts vanish completely and are replaced instead with aliasing, or half-frame displaced ghosts. The images are sharp, but any remnant phase error not accounted for leads to an interpolation error that manifests itself as ghosting rather than blurring.

There is therefore often more than one way to attempt to collect the same data, and the method of choice may be determined by some physical principles related to how the data are collected or to the type of artifacts. For example, for the shortest possible echoes, the asymmetric data from predominantly one-sided data collection would be the best choice, whereas if field inhomogeneities cause unacceptable artifacts, the method described in the preceding paragraph is the preferred one.

Collecting Multiple Echoes Adding a little more physics to the problem, one can ask the question, How can I supplement the lack of information available from, say, a very short asymmetric echo? One possibility is to add a second echo after this that may or may not itself be symmetric. The second echo will have worse phase errors than the first but the phases of the two are simply related. However, one must then deal with another reason why it is so difficult to carry out partial Fourier reconstructions: the data $s(k)$ are the Fourier transform of an object that is spatially varying, hence, all points in $s(k)$ contain information about all points in $\rho(x)$; deconvolving this information is computer intensive. It is possible to use this second echo along with a POCS-like iterative approach to more accurately solve this problem.

Two- and Three-Dimensional Extensions A largely ignored problem is the application of these data extension methods to two and three dimensions. A tremendous increase in time savings is possible in three-dimensional imaging if sampling is reduced in both phase encoding and slice encoding directions. The key issue is whether there is sufficient information available in any data acquisition scheme to uniquely solve the problem.

4.3.2 Reduced Gibbs Ringing

In the presence of a discontinuity, the Fourier transform yields an estimate that has a “ringing” at the discontinuity. This is true for any number of sampled points, and the Gibbs ringing is essentially scale invariant. This means that the ringing is always present at the same pixel values with similar amplitude. However, as the number of points in an MR image of given size increases, so also does the resolution, and so the points at which the ringing occurs are squeezed spatially closer to the discontinuity and thus become less likely to contribute to misinterpretation of the image. Recall that this famous Gibbs ringing or truncation artifact decays very slowly (as $1/x$ in the image domain), and so it is considered a particularly nasty artifact, one that can obscure information or mimic structure that is not really present physically. Therefore, quite some effort has been made in the MRI signal processing world to eliminate or reduce these effects.

Iterative Sigma Filtering One approach to reducing Gibbs ringing is to apply a nonlinear filter to determine the major edges (or discontinuities) in an image and then apply a sigma filter (or some other similar nonlinear filter) to smooth the image outside the edge. Interpolating the image by a factor of two, for example, and reconstructing a new data set with $4n$ points, leads to an extrapolated k-space data set that supposedly has recovered some of the missing high-k-space data points. This process can be repeated iteratively by modifying the filter values as a function of iteration number.

Constraint-based Methods Another approach to filtering images can be seen with MR angiographic images. Here the starting assumption is that the image can be converted to binary format: pixels can be segregated into two classes, those that depict parts of blood vessels and all others. Then resolution can be enhanced by taking partial volume effects into account to relabel some pixels. This method of filtering has been tackled rather nicely using a Bayesian approach, and more can perhaps be done with multiple image components.

Parametric Estimation Parametric estimation techniques are the most powerful methods for extracting information from an image when a priori information is available. Various approaches use linear prediction techniques and autoregressive moving average (ARMA) methods, localized polynomial approximation (LPA), and the generalized series method. These methods all attempt to recover parametric variables that describe a model. In the simplest one-dimensional example, modeling the data as a box, one would find the location of the center of the box, its amplitude, and its width. Unlike FFT methods or image processing methods, these methods have no inherent pixel size limitations. These are “super-resolution methods” that can be used to speed up image acquisition. For a box only two data points are needed!

What are the limitations of these methods and why are they not more commonly used? First, the resolution is now dependent on SNR. Second, they are often ill-posed (i.e., the condition number is large). Third, they are much slower to implement on a computer (often hundreds of times slower than the FFT). Nevertheless, with the dramatic improvement in computation speeds available today, the reconstruction times have dropped to within a few seconds in some cases, making these methods more viable in practice.

4.3.3 High-speed K-space Coverage Techniques

Moving from one-dimensional problems, one method in MRI that acquires data in a two-dimensional format and in a single shot is echo planar imaging (EPI). In fact one can cover k-space in a plethora of ways with rectilinear coverage, or in sinusoidal fashion. Each method has its own artifacts; sinusoidal coverage gives ghosts, while spiral scanning gives blurring. While speeding up imaging by simply taking fewer data points can always be done, but at the expense of resolution, these various methods are important because they attempt to overcome the resolution limits. In keyhole imaging, for example, a limited number of data are recollected as, say, a subject performs a functional imaging task where only a small portion of the image changes. Concepts such as use of shared echoes between scans to speed up spin echo or cardiac cine-imaging are similar to those described in section 4.3.1, but different enough to warrant rethinking certain data acquisition schemes that might otherwise not qualify as real-time imaging.

A final comment in this section relates to the general reconstruction question in n dimensions. First, what is a sufficient number of points to collect for a given reconstruction method? For example, for the FFT a high enough density of points can guarantee no aliasing even with non-uniform sampling. Second, how can we know how k-space is covered and, once we do, how can we correct for any deviations from the sampling we expected that are due to eddy currents or other phenomena? The first part of this question may be addressed using some features of the reconstruction method itself (such as varying the parameters until ghosting vanishes), but it is more likely answered by first collecting data that can serve as a calibration scan to determine background phase or, better yet, the exact location of all points to be sampled in k-space.

The reader should be aware that MRI is an incredibly flexible imaging modality. More often than not, thanks to the linearity of the Fourier transform and therefore of the whole data acquisition approach, many of the types of methods outlined here can be combined! By itself any one method may be unexciting; however, if each method were to give just a $\sqrt{2}$ improvement in SNR or speed, then several combined together could yield significant improvements overall.

4.3.4 Research Opportunities in Dynamic MR Image Reconstruction

Although many areas of image reconstruction have been examined as possible improvements to the FFT, few have succeeded in becoming practical enough for application in MRI. However, given ever-increasing computation speeds, the following topics may play a key role in the future of real-time scanning:

- Maximum entropy methods, although research to date has not shown them to be useful in MRI;
- Wavelets, for modifying both how data are collected and how reconstruction is performed;
- Bayesian analysis to improve signal-to-noise ratios; and
- New methods of iteration based on the search-everywhere concept of neural networks.

In all cases it is important to use the FFT as a benchmark and to ensure that the methods are stable and representative of reality under all reasonable imaging conditions for the modality.

4.3.5 Suggested Reading Related to Dynamic MR Image Reconstruction

1. Constable, R.T., and Henkelman, R.M., Data extrapolation for truncation artifact removal, *Magn. Reson. Med.* **17** (1991), 108–118.
2. Haacke, E.M., Liang, Z.-P., and Izen, S.H., Constrained reconstruction, a super-resolution, optimal SNR alternative to the FFT, *Med. Phys.* **16** (1989), 388–397.
3. Liang, Z.-P., Boada, F.E., Constable, R.T., et al., Constrained reconstruction methods in MR imaging, *Reviews of Magnetic Resonance in Med.* **4** (1992), 67–185.
4. Liang, Z.-P., and Lauterbur, P.C., An efficient method for dynamic magnetic resonance imaging, *IEEE Trans. Med. Imaging* **13** (1994), 677–686.
5. van Vaals, J., Brummer, M.E., Dixon, W.T., et al., “Keyhole” method for accelerating imaging of contrast agent uptake, *J. Magn. Reson. Imaging* **3** (1993), 671–675.
6. Wu, Z., Chung, H., and Wehrli, F.W., A Bayesian approach to subvoxel tissue classification in NMR microscopic images of trabecular bone, *Magn. Reson. Med.* **31** (1994), 302–308.

4.4 Applications of Dynamic MRI

4.4.1 Blood Flow

MRI has unique potential for imaging and quantifying blood flow. Flowing spins modulate the MR signal in two different ways. First, displacement of the bolus between successive excitations introduces some fully relaxed spins and thus an enhancement of the signal from the flowing blood relative to that of the stationary tissue. This time-of-flight effect is the basis of a major class of angiographic imaging techniques. The second effect has its origin in phase shifts imparted by the imaging gradients or specially administered flow-encoding gradients. This approach, summarized under the term “phase-contrast” (PC) MRI, is suited to the generation of angiographic images as well as quantitative velocimetry and rate measurements, or even to such applications as measurement of vascular compliance.

Measurement of Blood Flow Velocity PC MRI procedures are sensitive to velocity induced phase shifts in transverse magnetization. In these procedures, data are acquired first with flow-encoding gradient pulses of one polarity, followed by flow-encoding gradient pulses of the opposite polarity. The detected signals are typically processed in one of two ways. In the first approach complex differences of the acquired data are taken. Signals from stationary tissue have no velocity induced phase shift and thus cancel upon subtraction. Signals from moving spins, however, add and are preserved. In the second method, phase differences are computed. Complex difference images typically provide excellent morphological information, but only qualitative velocity information since the relationship between image pixel intensity and velocity is not linear. The phase difference images, on the other hand, provide a more quantitative presentation of velocity data because the phase shift of each pixel is directly proportional to velocity.

A significant practical shortcoming of the phase based methods is aliasing, which relates to the dynamic range of the image. The optimal SNR situation pertains when the highest measured velocity generates a phase that gives the maximum value on the gray scale image. In practice, however, the maximum velocity is not known a priori. The operator must then either choose a velocity-encoding level that encompasses a significantly larger velocity range and pay the penalty of a diminished SNR, or choose a lower value and run the risk that the maximum velocity exceeds the assigned dynamic range, in which case the measured velocity will be aliased and will appear as flow in the opposite direction. Schemes have been proposed to correct for this type of aliasing artifact, but such schemes must be used with caution to differentiate artifactual aliasing from hemodynamic effects.

Phase mapping methods have been used extensively for non-uniform (e.g., arterial) flow evaluations, which are typically performed using cardiac triggering with the acquisition of multiple phases in the cardiac cycle. Because the duration of the pulsatile cycle varies, many researchers choose to cover two cardiac cycles to ensure that velocities over the full pulsatile cycle are acquired, particularly when the aim is to measure volume flow. Alternatively, retrospective gating methods can be used. Pulsatile flow presents an added challenge since the velocity-encoding level must be chosen to encompass the peak velocity, resulting in reduced SNR in the remainder of the cardiac cycle. Schemes have been proposed to vary the velocity sensitivity through the pulsatile cycle.

Alternative spatial encoding strategies have been proposed to reduce image acquisition time. Both EPI methods and spiral scan methods have been investigated. These methods substantially reduce the time needed to acquire two-dimensional spatial information, requiring only “single-shot” excitation pulses rather than a number equal to the number of phase-encoding steps.

Fourier Velocity Encoding Phase mapping methods yield data in which the mean velocity in each image pixel is measured. Magnetization phase information can also be used to determine the velocity distribution within each pixel. This is accomplished by incorporating a cycle of phase-encoding steps (using bipolar gradients) designed to provide a phase shift proportional to spin velocity rather than spin position. An image can then be constructed in which one dimension of the image is a spatial dimension, one is a velocity dimension, and the distribution of velocities within each pixel is displayed. This method can be extended to encoding multiple spatial and/or velocity dimensions at the expense of additional scan time.

RF Pulses A key ingredient in the rapid advance made in both phase mapping and Fourier velocity-encoding methods is the use of two-dimensional selective excitation pulses. These pulses speed up data acquisition because they reduce the dimensionality of the object that must be encoded. They also reduce partial voluming and vessel overlap. Other RF pulse strategies are also attractive. Excitation by a “comb” of frequencies has been tried, which permits the simultaneous excitation of multiple and spaced parallel planes transverse to a vessel of interest.

Measurement of Wave Speed and Distensibility Increases in arterial pressure during the cardiac cycle arise from the contraction of the heart and are propagated to the vascular tree. If blood vessels were perfectly rigid, the pressure wave would be propagated at approximately the speed of sound in water. However, blood vessel walls are far from rigid, and pressure wave propagation is considerably slowed by the expansion of the vessel in response to the pressure wave. The degree of a vessel wall’s flexibility is frequently referred to as its distensibility D , which can be expressed as

$$D = \frac{\Delta A}{A_0 \Delta P}, \quad (4.7)$$

where ΔA is the change in cross-sectional area of a vessel in response to the pressure wave, A_0 is the vessel’s cross-sectional area at diastolic pressure, and ΔP is the pressure change during the cardiac cycle. Unfortunately, this definition of distensibility is not particularly useful with MR because of the difficulties in accurately measuring ΔA and ΔP .

An alternative method for the determination of distensibility is to measure the wave speed within the vessel. Distensibility is related to wave speed C by the relationship

$$D = \frac{1}{\rho C^2}, \quad (4.8)$$

where ρ is the density of blood.

Wave speed can be measured directly by observing the time delay of the pressure front at two stations a known distance apart. The accuracy of such a wave speed measurement is limited only by the accuracy of the distance and time measurement, and by the assumptions that (1) reflections of the pressure wave are absent and (2) the wave speed is substantially greater than the blood velocity. For many vessels these assumptions can safely be made. Consequently, detection of the onset of blood motion at each station is usually adequate.

Postprocessing As in other fields of MRI, rapid postprocessing methods are essential to permit velocity-encoding methods to be clinically useful. Conventional two-dimensional phase mapping methods require the subtraction of multiple-image data sets and the interpretation of the data in those sets. Current implementations require several minutes of data manipulation, thus preventing the use of additional scans to elucidate questionable areas.

The data contained in a velocity study can be copious and can have multiple spatial and velocity dimensions. The use of color coding, familiar to physicians from Doppler ultrasound, has been demonstrated and is potentially helpful.

Conclusions Related to MR Imaging of Blood Flow Magnetic resonance imaging can be used in a number of ways to obtain non-invasive quantitative measurements of a variety of physiological flow parameters. MRI can be used to measure aspects of blood velocity vectors in a vessel, including their spatial distribution, direction, and magnitude. Constraints on imaging time and data set size prevent the acquisition of images in which data are obtained with high resolution in three spatial dimensions, three velocity dimensions, and the temporal dimension (i.e., the cardiac cycle).

Nevertheless, useful data can be obtained in three of these dimensions with reasonable scan times. Two such applications are presented here. In the first, two spatial dimensions are obtained over the cardiac cycle. In the second, a velocity dimension is obtained with a single spatial dimension over the cardiac cycle. Additional spatial information is obtained by multiplexing the data from several stations along the vessel with a comb excitation.

Although MR blood flow imaging is coming into wide use, only a few physiological applications of MR flow measurement have been demonstrated to date. Many more exist and are likely to become important in the future.

For example, MRI can be used to monitor the response of vessel wall compliance and flow dynamics to pharmacological intervention. These MRI procedures would be useful for monitoring both fast-acting agents (such as nicotine) and slow-acting agents (such as cholesterol). Other potential applications include the correlation of changes in blood vessel morphology with changes in flow patterns, the quantitative measurement of cardiac parameters, and the analysis of model systems to advance understanding of fundamental blood flow physiology.

4.4.2 Diffusion Imaging

The effect of molecular diffusion—random displacements on a molecular scale—has been studied via nuclear magnetic resonance since the 1950s. More recently, there has been increased interest in measuring diffusion coefficients in tissue because it has been recognized that pathological processes such as stroke uniquely alter the diffusion coefficients; therefore, changes in diffusion coefficients can provide a new means for diagnosing conditions such as stroke damage. The measurement of diffusion is based on the irreversible dephasing spins experienced in the presence of magnetic field gradients.

Measurement of Diffusion Coefficients in vivo The coupling of MRI with measurements of diffusion has been developed mainly during the past decade. Despite the intrinsic sensitivity of diffusion MRI to motion artifacts, it has been demonstrated that, with care in the experimental settings and proper hardware, water molecular diffusion can be measured in vivo with MRI with fairly good accuracy and reproducibility, especially in the brain. Motion artifacts can be significantly reduced using navigator echoes or, more efficiently, EPI. With EPI, multiple images can be acquired in short time intervals, allowing greater accuracy in the measurements. Also, diffusion measurements can now be achieved in organs other than the brain, such as the kidneys or the heart.

Mapping of Diffusion Tensor Diffusion measurements may provide useful information on tissue microstructure and function, at a scale significantly smaller than the voxel size. Recently, it has been shown that the entire diffusion tensor can be determined in vivo, including in the human brain. The interest in diffusion tensor measurements, instead of simple coefficient measurements, has been spurred by the discovery that in many tissues, such as brain white matter and muscle, diffusion is highly anisotropic. This anisotropy is related to the tissue microstructure (muscle fibers, white matter fibers), which is also anisotropic, but the exact mechanisms are still unknown.

Diffusion intrinsically depends on temperature. Therefore, temperature images, or more exactly images of temperature changes, can be obtained with fairly good accuracy while the temperature of the object changes. This feature is particularly useful for monitoring interventional procedures performed within the magnet (“interventional MRI”).

4.4.3 Other Tissue Parameters

One of the main strengths of MRI is that the image intensity can be made to vary significantly depending on the type of acquisition sequence. This richness of image contrast parameterization leads to many applications in which quantitative assessment of various physiological or biophysical parameters is used to distinguish disease processes. How best to use the multi-parametric nature of the imaging data to maximize diagnostic information is a research area that has by no means been exhausted. Several examples of imaging parameters that provide more than simply anatomical mapping are discussed in this section.

Relaxation Times The contrast in conventional MR images depends on the nuclear magnetic resonance relaxation times of tissues, particularly the longitudinal relaxation time T_1 and the transverse relaxation time T_2 . Other relaxation times, such as T_2^* , or specific relaxation mechanisms, such as magnetization transfer, can also be determined. For each of these parameters, there are image acquisition sequences that allow direct spatial mapping of this information. Several parameters can be used in concert for maximum discrimination of tissue types as a basis for image segmentation. Quantitation of relaxation times as a means of distinguishing pathologies has not received wide acceptance, perhaps because of intrinsic biological heterogeneity or, more likely, unreliable measurement techniques for these relaxation times. However, quantitative relaxation times provide information about tissue pathology in several instances: trabecular structure, recurrent cervical carcinoma versus post-radiation fibrosis, and muscle response of exercise have all been measured with quantitative relaxation times. Relaxation times in tissues have also been shown to represent multi-component decays. This adds to the wealth of quantitative capability (and also to the unreliability of many single-component measures). Understanding the biophysical source of these various components remains a challenge, and combining them appropriately to obtain maximum diagnostic information is an ongoing research opportunity.

Oxygen Of specific clinical interest is the measurement of the oxygen provided to tissues. MRI can provide oxygen-dependent signals in the major vessels due to T_2^* differences for oxygenated blood versus oxygen-deficient blood. This is one of the mechanisms that contributes to the functional brain MR effect. Quantitation of these differences in the presence of flow presents technical challenges. More importantly, direct measures of oxygen in tissue may be achievable by looking at myoglobin or at oxygen-dependent measurements from fluorine in fluorinated hydrocarbon blood substitutes.

Strain Particularly for cardiac functional imaging, measures of cardiac strain are important indicators of heart wall function. MRI can provide such measures either using direct velocity encoding, as is done in blood flow, or using tagging techniques that lay a grid on the heart image and then look at the subsequent deformation of the grid at a later time. Both of these techniques are able to yield the basic data for strain mapping. Since strain is a tensor quantity, the eventual extraction and display of strain maps constitute a computer graphics challenge, although significant progress is being made on this frontier. A full and detailed map of cardiac strain would provide much of the information needed for a complete cardiac diagnosis. From a logistical point of view, this capability could make MRI a primary diagnostic modality for cardiac evaluation.

Very recently, it has been shown that the phase mapping methods discussed in section 4.1.1 for studying fluid dynamics can be applied to measuring the response of body tissue to harmonic mechanical excitations. For example, synchronizing an oscillating motion-encoding gradient with the actuator that induces the mechanical wave results in phase shifts in the images from which displacement maps are obtained, enabling the computation of elastic moduli.

4.4.4 Functional Brain MRI

Since its inception in 1973 MRI has evolved into one of the most powerful non-invasive techniques in diagnostic clinical medicine and biomedical research. However, MRI is used primarily as a technique for producing anatomical images. A very significant recent development, suggested by Seiji Ogawa of AT&T Bell Laboratories and applied to humans simultaneously at the University of Minnesota, the Medical College of Wisconsin, and Massachusetts General Hospital, has been the use of MRI to non-invasively map human cortical function. This new technique has been dubbed “functional magnetic resonance imaging,” or fMRI, and was almost immediately reproduced by other workers in the field, using both conventional gradient echo imaging and EPI.

Contrast Mechanism The basis of fMRI lies in the fact that deoxyhemoglobin, found in red blood cells, acts as nature’s own intravascular paramagnetic contrast agent. Deoxyhemoglobin, that is, hemoglobin without a bound oxygen molecule, is a paramagnetic substance. When placed in a magnetic field, a blood vessel containing deoxyhemoglobin alters the magnetic field in its vicinity. The effect increases as the concentration of deoxyhemoglobin increases, and at concentrations found in venous blood vessels a detectable local distortion of the magnetic field surrounding the vessel is produced. This distortion can affect the MR behavior of the water protons within and surrounding the vessels, an effect that manifests itself as a small but detectable change in the image intensity of appropriately acquired MR images.

Increases in neural activation within the cerebral cortex (such as are brought about by a task, stimulus, or seizure) lead to an increase in blood flow without a commensurate increase in oxygen extraction. Hence, the capillary and venous deoxyhemoglobin concentrations decrease as oxygen-rich arterial flow increases with no increased oxygen extraction. The decreased concentrations are reflected in increases in the MRI relaxation times T_2^* and T_2 and consequent increases in the signal intensity of T_2^* - and T_2 -weighted MR images (the so-called blood oxygen level dependent, or BOLD, effect). Functional MRI has been successfully applied to delineate the activity in the human visual cortex using exogenous contrast agents or endogenous BOLD. The BOLD approach has already been extended to the examination of other somatosensory and cognitive tasks, such as motor movement and speech.

Imaging Techniques The further development of imaging techniques and devices will be a fruitful area for research in fMRI. For the reasons mentioned above, high-speed imaging methods are needed that are capable of covering the entire brain, to follow either an exogenous tracer or the endogenous changes that occur during activation.

A variety of techniques are capable of imaging the brain at different time scales. Echo-planar imaging (EPI) typically acquires images on the 20- to 100-ms domain. There are several fast “conventional” gradient echo techniques that image in the 1- to 60-second domain, as well as a variety of novel sequences that combine elements of both techniques, including spiral imaging and sequences known by the acronyms DUFIS and BURST. The areas of research are outlined below.

EPI was first described by Mansfield but has only recently been available commercially. In EPI, the entire MR image is encoded following a single RF excitation. The resulting imaging and reconstruction process can have a point-spread function, eddy current effects, and distortions that are substantially different from those of conventional MRI. In addition, while it is readily implementable on small-bore imaging systems, modifications to the imaging

system are required for human studies. These modifications, which remain an active area of investigation, include the use of resonant gradient systems, dedicated head gradient coils, and modified high-power linear systems. The implementation and use of modified EPI sequences on conventional scanners also represent an active area of research.

A growing set of alternatives to EPI offer different sets of advantages and are also applicable to fMRI. Spiral scanning techniques, for example, offer improved flow behavior. Many aspects of spirals present challenges, including both the efficient reconstruction and novel amplifier design for efficient implementation on whole-body systems. While both spirals and EPI attempt to cover k-space rapidly using rapidly changing gradients, other techniques, including the aforementioned BURST and a method called OUFIS, perform this coverage using rapid RF-encoding techniques. These techniques may allow fMRI without modification of hardware or the reconstruction software. The implementation and imaging issues (artifacts, point-spread function, and so on) are being actively pursued.

Finally, conventional MRI techniques are being adapted for fMRI of both exogenous and endogenous tracers. Modification of conventional techniques that improve the contrast/time relationship is important. These areas of research are particularly attractive, given the large installed base of conventional imaging systems on which such sequences would be available. Understanding and correcting motion artifacts are critical considering the small changes observed in fMRI.

Hardware Requirements Since EPI requires very fast alternating currents through the gradient coil, development of gradient coils that are torque-free, produce less acoustic noise, and/or are tailored for certain anatomic regions is currently under investigation. Also important are the bioeffects of the large dB/dt produced by these coils.

In terms of the imaging hardware as well, the requirements for successful fMRI are more stringent than those for conventional MRI. The data acquisition system must be highly synchronized, so that all pulses and data acquisition begin at precisely known (± 100 ns, or better) times on each shot. This temporal coherence must be maintained for the entire length of a data acquisition run (minutes); otherwise, the interimage fluctuations add to the noise level and make brain activity harder to detect. It is important to have an fMRI data quality assurance plan operating at any facility carrying out studies over a long period of time. Problems with the scanning system that may not greatly affect normal clinical images need to be detected and corrected, since they may degrade the fMRI results much more.

Field Strength Considerations As discussed above, local field gradients induced by differences in magnetic susceptibility between the intra- and extravascular spaces induce intravoxel dephasing of water protons and result in signal contrast in the image. This intravoxel phase dispersion directly contributes to a decrease in the apparent transverse relaxation rate T_2^* . If water diffusion distances are large compared to the spatial variation in these local gradients, some of the intravoxel phase dispersion is averaged and, as a consequence, signal loss due to dephasing is diminished. Although the dephasing effect is mitigated by the diffusion-related averaging, diffusion in the presence of the field gradient is a T_2 relaxation mechanism and can lead to a decrease in T_2 .

While the experimental demonstration of these effects is relatively trivial in terms of the pulse sequences required, a quantitative understanding is much more difficult to achieve. It is possible, however, to perform calculations that yield semi-quantitative predictions that can then be experimentally tested to increase understanding of the phenomenon. Based on such calculations, it is possible to predict that the field dependence of the alteration in T_2^* due to susceptibility-induced gradients around blood vessels will depend on the diffusional averaging; as such, the change in T_2^* will be a function of the motion of the protons in the brain tissue as well as the magnitude of the field gradients encountered by the tissue protons. The latter is, in turn, dependent on the size of the blood vessels and their orientation relative to the main magnetic field direction.

For a gradient-recalled echo that is sensitive to T_2^* , the signal will have between a linear and a quadratic dependence on magnetic field. These predictions are supported by data from experiments in which the signal intensity change due to visual stimulation was examined in the same individual in the same region of the brain at both 1.5 T and 4 T.

Thus, with respect to signal intensity changes (ΔS) induced by alterations in the susceptibility of blood vessels secondary to neuronal activity, the effect will increase more than linearly with magnetic field strength. However, this relationship does not guarantee that the higher field is superior for functional imaging. In order to address this question, one must consider $\Delta S/N$ where N is noise defined as the fluctuations in the signal intensity of the images obtained consecutively during the functional imaging paradigm. Contributions to this noise can come from instrument instability and physiological variables, as well as from the inherent noise in a given image. It has been shown that $\Delta S/N$ is maximum if the echo time T_E is chosen such that $T_E = T_2^*$, in which case

$$\Delta S/N \propto (S_0/N)(\Delta R_2^*/R_2^*), \quad (4.9)$$

where S_0/N is the SNR at $T_E = 0$, and $\Delta R_2^* \equiv 1/T_2^*$ and R_2^* are the relaxation rates under baseline conditions and the change in $1/T_2^*$ due to neuronal activation, respectively. Each of the quantities in equation 4.9 has a potential field dependence. Each of the four quantities in this equation is known to some extent. However, the exact nature of this relationship is currently not known and is a field of substantial interest and ongoing research. Clearly, if field

strengths in excess of 1.5 T were shown to be of significant benefit to functional imaging, there would be increased interest in the development of high-field imaging instrumentation.

Processing of Functional Images In order for fMRI to become a practical neurological research and clinical investigation tool, a large number of technical issues need to be addressed, including:

1. Motion detection and compensation;
2. Characterization of temporal response patterns;
3. Characterization of physiological noise and its effects on fMRI; and
4. Display of volume functional MR images, especially in real time.

Motion Detection and Compensation: Subject motion (relative to the image-defining gradient coils) during a scanning session is a vexing problem. As brighter and darker regions move in and out of a voxel, the intensity level of that voxel will fluctuate in time, regardless of any other physiological activity. Motion coherent with the task/stimulus timing will mimic functional MR activation; motion not coherent with the task/stimulus timing will add to the noise level and make the detection of true activations more difficult. This problem is the greatest where the tissue MR signal gradient is the largest—for example, at the interface of gray matter and cerebral spinal fluid, which is also where neuronal activation occurs. Detecting and compensating for motion effects is one of the most urgent needs in functional MR imaging.

Characterization of Temporal Response Patterns: The fMRI activation signal has been shown to vary across the brain. Whether this variance represents a physiological difference in the hemodynamics of different cortical regions or a difference in the rate at which different neurons are recruited for task performance is as yet unknown. One current method for detecting activation from the functional MR time series of images is to form the correlation of each voxel's intensity time series with an ideal response time series; large correlations are deemed to signify “active” voxels. The variability in the brain's response poses questions as to which ideal function to use, how important the choice is, and how to adapt the detection methodology to allow for the different response patterns.

The correlation method can be thought of as a least squares fitting of the ideal $r(t)$ to the observed $x(t)$ via the two parameter model $x(t) \approx \alpha r(t) + \beta$, where α represents the activation magnitude and β is the mean signal level in the voxel from which $x(t)$ is drawn. Expressed in this manner, an obvious question is, How many useful parameters can be estimated from an fMRI time series? Should linear signal models be used or can nonlinear models more economically and robustly represent the patterns of signals that are seen?

Another issue is variability in response between different iterations of a mental task or stimulus. This will tend to reduce the goodness-of-fit to any low-dimensional model (and high-dimensional ideal response models are not likely to be robust). Can such changes be better allowed for using a data driven approach (e.g., principal components analysis) to extract the ideal response signal model? Can inter-iteration variability be reliably quantified and used to judge task performance and/or learning?

A related question is the characterization and detection of responses that are not “locked” to any a priori known time(s). Not all desirable neurologic investigations fit neatly into the “task A:task B” type of alternation that is the prevalent mode at present. Can fMRI be used to determine when certain neural processes are taking place, as well as where?

Characterization of Physiological Noise and Its Effects on fMRI: The dominant “noise” source in fMRI is not thermal noise in the MR receiver coils or eddy currents in the sample volume. Physiological fluctuations are several times larger and have a complex spatio-temporal structure. The causes of this “physiological noise” may include pulsing motions caused by the heartbeat, oxygenation level changes caused by respiration, and local blood flow variations on the scale of 10–20 s. To improve the techniques now used, it will be necessary to gain a physical and statistical understanding of this physiological noise.

In summary, the range of issues related to motion detection, temporal response, and physiological noise must separately and jointly be studied in a systematic fashion, probably through statistical methods. Functional MR image processing can then be adapted to extract the maximum information from the huge quantity of raw data that is generated by an fMRI scanning session.

Display of Volume Functional MR Images, Especially in Real Time: To be really useful, fMRI must be done in three dimensions, either with multislice scans or with true volume imaging. Displaying the results in a useful manner is not a trivial task, especially in real time, where the results are changing with every new acquisition. At present, two-dimensional color overlays of activated regions onto high-resolution anatomical images are the display method of choice in published papers. This technique is easily adaptable to real-time fMRI, and provides a shot-by-shot display of brain regions that are above the activation threshold. Static three-dimensional display of fMRI results is difficult enough, and no standard method has yet been settled on. Three-dimensional color rendering in real time is

certainly practical on high-end workstations now (and will become more common in the near future), but the utility of a three-dimensional rendering that is updated every few seconds is not yet clear.

An issue of special concern to EPI is the registration of the functional MR images to the high-resolution anatomical images. Distortions in the magnetic field B_0 cause distortions in the echo-planar images that are not present in the anatomical reference images (typically gathered with a fast spin echo technique). Another registration issue is the projection of fMRI activation results onto the cortical surface. One potential medical application of fMRI is brain functional mapping to aid in neurosurgical planning. For the MR-mapped functional foci to be useful to a surgeon, they must be identifiable in the surgical field. One possible approach to this problem is registration of vessels detected by MR angiography (MRA) to make surface veins visible once the craniotomy has been performed. Considering the time constraints of surgery, this would have to be done very quickly after the patient's brain has been exposed.

Safety Considerations Important to the continued development of fMRI techniques, both at high speeds and at high field, is an understanding of various bioeffects produced by the imaging. Both empirical and theoretical modeling of such effects is actively being pursued. These efforts include the measurement and modeling of nerve stimulation due to rapidly changing magnetic fields (dB/dt effect), local deposition of RF energy due to eddies produced in the body, and other effects of high magnetic field strength. Of particular relevance would be more complete models (e.g., realistic finite-element models) that include the known inhomogeneities of both the biology and the imaging systems.

Biophysical Modeling In addition to studies in the physics and engineering of imaging systems, there are research challenges in the physics and mathematics of understanding and quantifying the underlying effects of fMRI as well as modeling, in humans, the imaging techniques used to elicit such data. Unlike x-ray or nuclear techniques, the changes observed in fMRI are produced by indirect mechanisms. Observed proton relaxation rates can be affected by compartmentalized paramagnetic contrast agents including deoxygenated hemoglobin, as well as by in-flowing blood. While the observation of these changes is robust, disentangling the multiple effects and their importance (e.g., vascular vs. parenchymal changes, arterial vs. venous changes, blood volume, blood flow, and oxygenation changes, glycolytic and non-glycolytic metabolic changes) is a prime focus of current research. The question of quantification also remains: Can combinations of MR images be used to unambiguously calculate changes, in physiological units, of oxygen consumption or of blood flow?

4.4.5 Multinuclear MRI

MR Spectroscopy and Spectroscopic Imaging Magnetic resonance has the unique capability of providing in vivo information on the chemical composition of human tissues, including brain, heart, muscle, liver, and other tissues, non-invasively through a combination of imaging (MRI) and spectroscopy (magnetic resonance spectroscopy, or MRS). Spatial and temporal variations in tissue function can be evaluated by examination of the spectra and spatial distribution of ^1H , ^{13}C , $^{14,15}\text{N}$, ^{19}F , ^{23}Na , ^{31}P , ^{39}K , or $^{35,37}\text{Cl}$. The method is known as spectroscopic imaging (SI) or chemical shift imaging (CSI). Many of the challenges relating to the implementation of SI are concerned with overcoming the limitations in data interpretation due to the relatively low SNR of the data and require the application of signal processing and parameter estimation techniques such as are discussed in Chapter 13. While recent improvements in MR hardware have made possible more flexible data acquisition, the design of time-efficient sampling strategies and the development of algorithms for data analysis are areas of ongoing research. Of particular interest to physics and mathematics research are strategies for incorporating prior information by exploiting knowledge of the spatial distribution or the spectral parameters of particular metabolites.

Whereas conventional MR images have good SNR due to the relatively high receptivity and natural abundance of protons, and the high concentration of water protons in tissues, other nuclei that can be detected by nuclear magnetic resonance are of much lower abundance and have lower receptivity (e.g., ^{23}Na has 13% of the sensitivity of protons, while ^{31}P has 8% and ^{13}C has only 0.025%). Furthermore, nuclei such as ^{23}Na , ^{39}K , and $^{35,37}\text{Cl}$ have nuclear electric quadrupole moments. Because these ions are bound to macromolecules, with their long rotational correlation times, the electric quadrupole moments are modulated by the local electric field gradients to result in very short relaxation times and thus broad lines. This presents technical challenges but offers opportunities as well, because difference in relaxation behavior inside and outside the cell permits differentiation between intra- and extracellular environments.

Techniques for Spectroscopic Imaging with ^{31}P : The metabolites that can be studied with ^{31}P SI include inorganic phosphate, phosphocreatine, adenosine triphosphate, phosphomonoesters (which include phosphocholine and phosphoethanolamine), and phosphodiester (which include glycerolphosphocholine and glycerolphosphoethanolamine). Localization is achieved through phase-encoding gradients alone or through a combination of volume selection and phase encoding. The phase encoding can be applied in one, two, or three dimensions, providing localization within slices, columns, or boxes. Most techniques currently use k-space encoding similar to that used for conventional MRI, followed by Fourier transforming to obtain the spatial information. Alternatively, volume selection may be achieved using a combination of gradient and RF pulses that excite either a slice, column, or box. Instead of obtaining two-

or three-dimensional arrays of intensities, as is the case with MRI, the SI data are Fourier transformed to produce a multi-dimensional array of spectra that correspond to a specific region of space. Although most studies have acquired data on a rectangular grid in k-space and reconstructed the spatial dependence of the data via a Fourier transform, alternative sampling strategies include Hadamard encoding and reconstruction and non-uniform k-space sampling. Other techniques assume prior knowledge of the spatial distribution of metabolites in order to reduce acquisition times.

Techniques for Spectroscopic Imaging with Water-Suppressed ^1H : The metabolites available for analysis by proton SI include lactate, N-acetyl aspartate, choline, creatine, GABA, glutamine, glutamate, and taurine. Other metabolites may be available depending on the SNR of the particular case.

The sequences that have been used for studying spatial variations in ^1H metabolites usually include water suppression, volume selection, and phase encoding. Water suppression is needed to attenuate the intense water resonance down to a level where it is possible to detect metabolites at concentrations 100 to 1000 times lower. Volume selection can be used to restrict the size of the region being studied or to reduce the contribution of spurious signals in the spectrum, such as the lipid resonances that correspond to subcutaneous fat. Two of the most commonly used techniques for acquiring ^1H SI data use volume selection in combination with one-, two-, or three-dimensional phase encoding. The relatively high sensitivity of ^1H SI has made it possible to obtain data from the brain at a spatial resolution of 1–2 cm³ with a volume coil and 0.15–0.3 cm³ with phased array or surface coils.

Other approaches to volume selection and water/fat suppression use spatially and frequency-selective pulses, outer volume saturation pulses, or selective inversion recovery. These techniques have been applied to obtain two-dimensional arrays of spectra from entire slices through the head with an acquisition time of about 30 minutes for three or four 15 mm slices, with a slice gap of 5 to 10 mm. While the multislice capability is useful in reducing acquisition times, the clinical evaluation of whole brain metabolism is still extremely demanding; it requires good shimming and fat suppression over the whole head, as well as acquisition of data from contiguous slices. The latter implies the use of interleaved acquisitions, which would double the acquisition time. This may prove to be feasible by means of high-speed SI techniques based on the principles of echo-planar imaging. These more-rapid data acquisition schemes demand more complex data reconstruction and thus offer opportunities for future research.

Processing and Analysis of SI Data: Experiments using two- or three-dimensional SI may produce hundreds of spectra in a single examination. Processing such data routinely would require substantial automation as well as the development of algorithms for quantifying spectral parameters in the presence of relatively low SNR. Approaches to solving these problems are discussed in Chapter 3. A further requirement for efficient interpretation of the data would be graphical tools for correlating spectra and images. Several approaches for relating metabolic and anatomical data have been proposed, including plotting arrays of spectra and superimposing corresponding grids on the MR images, making tables of the distribution of each metabolite, overlaying full MRI or edge-detected MRI data with metabolite images, and using segmentation (see section 13.1) to classify morphological tissue types in order to estimate the fractions of each tissue type corresponding to spectral voxels. For qualitative interpretation of the data, the image format is appealing but, for the experienced observer, spectral arrays are more valuable in judging the overall quality of the data, determining the metabolite levels in a specific region, and relating information about different metabolites. Thus, it is not clear at this stage which approach is best suited for visual interpretation of clinical studies and how the spectral information will ultimately be best integrated with the imaging data.

Injected Paramagnetic Contrast Agents and Hyperpolarized Noble Gases Although this report does not give emphasis to the active field of tracer chemistry as it applies to MRI, positron emission tomography (PET), and single photon emission computed tomography (SPECT), or to the mathematical aspects of the biodistribution of the agents in the body, a few important fundamental principles of biodistribution should be mentioned, along with some new developments that bear on future applications.

Injected paramagnetic contrast agents such as gadolinium or dysprosium chelates are used with MRI to track the spatial distribution over time of the agent, from which kinetic parameters such as flow or tissue perfusion can be derived using mathematical models of various degrees of complexity. These agents effect a change in the local relaxation rate $1/T_1$ that is proportional to the concentration of the contrast agent. The paramagnetic contrast agent is bound to a radionuclide tracer so that the concentration of the contrast agent in the organ or region of interest can be monitored. The sensitivity of MRI is too low for detection of the T_1 changes associated with neuroreceptor-targeted ligands or antigens in general.

Another class of contrast agents appropriate for MRI medical applications is the noble gases with spin 1/2, notably hyperpolarized ^3He and ^{129}Xe . As opposed to the situation with paramagnetic contrast agents, these hyperpolarized gases serve as both tracer and the source of the MRI signal. While the sensitivity of MRI is too low to detect the normal concentrations of injected nuclei, because only a few nuclei out of one million would be polarized, 10 to 20% of the nuclei are polarized with these gases, and the sensitivity increase of 10,000 to 100,000 enables their imaging. The hyperpolarized gases maintain their polarization for many minutes, or even hours if bottled in a magnetic field and stored at cryogenic temperatures.

These inert gases can be used for lung imaging and, because of their finite solubility in tissue, could potentially be used for tissue perfusion quantitation as has been done with radioactive xenon, ^{127}Xe and ^{133}Xe . Alternatively, one can utilize a known signal decay of the nuclear spin relaxation rate, $1/T_1$, to determine flow at equilibrium during constant infusion, probably through inhalation of a dilute mixture of the hyperpolarized gases.

4.4.6 Microscopic Imaging

There is no well-defined boundary between standard macroscopic imaging and microscopic imaging. A common working definition places that boundary at the best resolution attainable with the usual clinical MRI systems, perhaps about $300\ \mu\text{m}$ for protons. The attainable resolution using surface coils can be less than $50\ \mu\text{m}$, so that MRI covers a volume range of about 10^6 , with a corresponding range of total numbers of nuclei and hence signal strength. The potential applications range from the improved delineation of anatomical structure already resolvable by MRI to new classes of features in tissues. Examples of specialized applications that may become important include visualization of cortical layers, the precise boundaries of functional regions in the brain, and microangiography at the level of arterioles and venules. All of these applications may be augmented, or made possible, by the use of magnetic contrast agents to allow the identification of specific objects, such as blood vessels and tissues of particular types.

Under various circumstances, resolution may be limited by the achievable SNR, digitization (during acquisition or processing), bulk motion, or molecular diffusion, as well as by other problems and artifacts typically encountered in MRI.

Resolution It is common in practice to confound true image resolution with digital resolution (the dimensions of a pixel or voxel) and to emphasize pixel resolution in a thick slice. These practical working definitions can give rise to disagreements when more rigorous analyses of instrument performance and image interpretation are carried out.

There is no fundamental physical limit to resolution in MRI. The limits encountered in practice are complicated functions of the features of the apparatus, acquisition and processing methods, and object characteristics. In a given experiment, the limits may be set by the SNR in a single voxel, by digital resolution, by intrinsic line widths (homogeneous or heterogeneous broadening, by T_2 or susceptibility effects, for example), by diffusional effects, or by local or bulk motions.

Signal-to-Noise Ratios SNR may be increased by using higher magnetic fields, improved RF coil designs (including cooled coils), smaller (external, intrusive, or implanted) coils, more efficient acquisition methods, and special processing methods. When the contrast-to-noise ratio or the T_1/T_2 ratio is important, contrast agents may increase the effectiveness (the contrast attainable in a fixed time). Small objects can now be imaged with voxel resolution in the range from $100\ \mu\text{m}^3$ to $1000\ \mu\text{m}^3$, using receiver coils of the order of millimeters in diameter at magnetic fields of several tesla. It can be anticipated that regions near surfaces or implanted coils of similar dimensions will be imaged in vitro at similar resolution (about $10\ \mu\text{m}$ isotropically) in the near future. Ultramicro coils (10- to $100\text{-}\mu\text{m}$ diameter, with integral preamplifier) may improve the volume resolution by about an order of magnitude because they increase the SNR by a similar factor. Because in the small-coil limit the noise arises predominantly from the coil rather than the sample, the recent development of superconducting receiver coils has afforded dramatic enhancements in sensitivity that, in turn, can be traded for improved resolution by decreasing voxel size.

Gradients As the SNR is increased, the gradient strengths required to encode image information become a problem. High-amplitude gradient coils to surround small objects are readily constructed and operated, but in vivo applications require special structures, efficient encoding methods to minimize the gradient amplitudes required, and low-duty cycles or effective cooling to avoid local heating. These considerations probably limit resolution to the order of a few tenths of a millimeter for the immediate future in human application.

Diffusion Molecular diffusion is becoming a useful source of contrast and microstructural information in studies at ordinary resolution. It plays the same role in microscopy, although unwanted diffusion damping may be a more important factor in microscopy because of the large gradients. As resolution of tens of micrometers is approached, diffusion displacements of nuclei become more important, and below a $10\text{-}\mu\text{m}$ pixel size molecular diffusion becomes the dominant consideration in experimental design.

Motion The effects of motion may be counteracted by faster image acquisition, gated acquisition (to the cardiac cycle, for example), and a variety of other tracking and processing methods already used in clinical practice and research. Postprocessing, such as realignment of images to compensate for rigid body motions, is becoming more practical as readily available computing power increases. Newer methods, such as (k, t) -space imaging, may provide a more general solution to the problem. Within some regions of the brain, motion amplitudes may be only a few hundred micrometers or less. Elsewhere in the body the problems will almost always be greater, and specific solutions may be required. It should be noted that the coordinate system is defined by the magnetic field gradients, not by

spatial location itself. Movement of gradient coils can be mistaken for object motion, and gradients that track object motion would eliminate primary motion artifacts, but are not now practical.

Future Applications of in vivo MRI Microscopy A rich variety of structural and functional features becomes visible below the millimeter level. Their visualization will usually be possible with apparatus and techniques developed for specific structures in specific locations, and for studies of a well-defined problem of function or diagnosis.

4.4.7 Research Opportunities Related to Applying Dynamic MRI

Blood Flow All MR velocity studies will benefit from hardware improvements that affect flow image quality. These include dedicated RF coils, high-performance magnetic field gradients, reduced eddy current effects, and improved spatial homogeneity, to name a few. Software improvements to enable spatially selective RF pulses are also vital. Postprocessing and display of velocity information are currently tedious and impractical. Improvements in this area must also permit the exploration of additional information in the data, such as the extraction of pressure values and shear/stress forces. Some specific research opportunities include the following:

- Development of techniques for rapid measurement of instantaneous velocity in three-dimensional space (six-dimensional problem);
- Modeling of complex flow and its implications for the vascular MR signal; and
- Development of methods for extracting parameters of physiologic relevance, including shear stress, distensibility, and turbulence.

Diffusion Imaging One field of research regards the implementation of diffusion tensor imaging (DTI) with EPI on conventional systems for clinical use. The implementation of EPI would require improvements in the gradient hardware (eddy current compensation and high-amplitude, high-slew-rate gradients), which would also benefit diffusion accuracy. Further, understanding the mechanisms underlying the diffusion values in tissues would be highly desirable. Most diffusion values are about one order of magnitude smaller in tissues than in pure water. Part of this difference can be explained by the tissue microstructure, in terms of obstacles, fibers, or membranes, but some of the differences may be artifacts. Matching MRI diffusion measurements by DTI with direct measurements using microelectrodes in tissues or animal preparations would be extremely useful. With those techniques the existence of restricted diffusion effects, requiring ultrashort diffusion times, could be demonstrated. Also the mechanisms of anisotropy in brain white matter could be better understood, as well as those involved in acute brain ischemia.

The clinical value of diffusion MRI in stroke needs more evaluation. What is the prognostic value of diffusion in terms of patient recovery? How would repeated diffusion measurements help monitor the effects of drug therapy? On the other hand, would anisotropic diffusion studies benefit management of patients with myelin disorders (e.g., brain development retardation, multiple sclerosis)? More generally, what would be the role of DTI in the evaluation of brain diseases? Still to be done are in vivo studies demonstrating the feasibility of the application of DTI to temperature imaging in interventional MRI, as well as evaluations of the effects of blood flow and tissue denaturation on the diffusion measurements.

Some specific research opportunities related to diffusion imaging include:

- Establishment and experimental validation of a molecular model for anisotropic diffusion in tissues;
- Development of improved approaches for spatially localized measurement of diffusion coefficients in vivo;
- Modeling of heat dissipation in tissue; and
- In vivo measurement of tissue fiber orientation.

Other Tissue Parameters

- Development of accurate measurement techniques for quantitative relaxation times and their interpretation in terms of clinical diagnosis.
- Development of sophisticated segmentation techniques based on multiple parametric acquisitions.
- Extraction and meaningful display of strain maps of cardiac function.

Functional Brain MRI

- Development of methods for monitoring data quality during a scanning session and for ensuring that functional activation is being observed.
- Development of new experimental protocols, especially for use with complex stimuli (e.g., visual presentations).
- Modeling and experimental verification of the biophysical contrast-to-noise mechanisms induced by neuronal activation and their dependence on magnetic field strength.
- Establishment of detailed biophysical models to understand the stimulus response permitting separation of vascular from parenchymal changes and arterial from venous changes, taking into account parameters such as blood volume, blood flow, and so on.
- Evaluation of the nature of various sources of noise (e.g., stochastic, physiological, instrument instability) and development of strategies for their minimization.

Multinuclear MRI

- Evaluation of polarized noble gases as tracers of pulmonary function and tissue perfusion (e.g., in muscle and in the brain and other organs).

Microscopic Imaging

- Exploration of the theoretical limit of spatial resolution and its dependence on key parameters, including diffusion and detection sensitivity.
- Development of improved means for monitoring of and correcting for the effects of motion, which currently limits resolution of in vivo MR microscopy.

4.4.8 Suggested Reading on Applications of Dynamic MRI**Blood Flow**

1. Bryant, D.J., Payne, J.A., Firmin, D.N., et al., Measurement of flow with NMR imaging using gradient pulse and phase difference technique, *J. Comput. Asst. Tomog.* **8** (1984), 588–593.
2. Caro, C.G., Pedley, T.J., Schroter, R.C., and Seed, W.A., *The Mechanics of Circulation*, Oxford University Press, Oxford, 1978.
3. van Dijk, P., Direct cardiac NMR imaging of heart wall and blood flow velocity, *J. Comput. Asst. Tomog.* **8** (1984), 429–436.
4. Dumoulin, C.L., Souza, S.P., Hardy, C.J., and Ash, S.A., Quantitative measurement of blood flow using cylindrically localized Fourier velocity encoding, *Magn. Reson. Med.* **21** (1991), 242–250.
5. Dumoulin, C.L., Souza, S.P., Walker, M.F., and Wagle, W., Three-dimensional phase contrast angiography, *Magn. Reson. Med.* **9** (1989), 139–149.
6. Feinberg, D.A., Crooks, L.E., Sheldon, P., Hoenninger III, J., Watts, J., and Arakawa, M., Magnetic resonance imaging the velocity vector components of fluid flow, *Magn. Reson. Med.* **2** (1985), 555–566.
7. Mohiaddin, R.H., Firmin, D.N., Underwood, S.R., et al., Aortic flow wave velocity: The effect of age and disease, *Magn. Reson. Imaging* **7** (suppl. 1) (1989), 119.
8. Wedeen, V.J., Meuli, R.A., Edelman, R.R., et al., Projective imaging of pulsatile flow with magnetic resonance, *Science* **230** (1985), 946–948.

Diffusion Imaging

9. Basser, P.J., Mattiello, J., and Le Bihan, D., Estimation of the effective self-diffusion tensor from the NMR spin-echo, *J. Magn. Reson., B* **103** (1994), 247–254.
10. Le Bihan, D. (ed.), *Diffusion and Perfusion Magnetic Resonance Imaging: Applications to Functional MRI*, Raven Press, New York, 1995.
11. Le Bihan, D., Molecular diffusion nuclear magnetic resonance imaging, *Magn. Reson. Q.* **7** (1991), 1–30.
12. Reese, T.G., Weisskoff, R.M., Smith, R.N., Rosen, B.R., Dinsmore, R.E., and Wedeen, V.J., Imaging myocardial fiber architecture in vivo with magnetic resonance, *Magn. Reson. Med.* **34** (1995), 786.
13. Moseley, M.E., Cohen, Y., Kucharczyk, J., et al., Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system, *Radiology* **176** (1990), 439–446.
14. Turner, R., Le Bihan, D., Maier, J., Vavrek, R., Hedgers, L.K., and Pekar, J., Echo-planar imaging of intravoxel incoherent motions, *Radiology* **177** (1990), 407–414.

Other Tissue Parameters

15. Hwang, S.N., and Wehrli, F.W., The calculation of the susceptibility-induced magnetic field from 3D NMR images with applications to trabecular bone, *J. Magn. Reson., B* **109** (1995), 126.
16. Kroeker, R.M., and Henkelman, R.M., Analysis of biological NMR relaxation data with continuous distributions of relaxation times, *J. Magn. Reson.* **69** (1986), 218.
17. Labadie, C., Lee, J.-H., Vetek, G., and Springer, C.S., Jr., Relaxographic imaging, *J. Magn. Reson., B* **105** (1994), 99–112.
18. Li, K.C.P., Wright, G.A., Pelc, L.R., Dalman, R.L., Brittain, J., Wegmueller, H., Lin, D., and Song, C., In vivo verification of MR measurements of superior mesenteric vein blood oxygen saturation in a canine model, *Radiology* **194** (1995), 321–326.
19. Moore, C.C., O'Dell, W.G., McVeigh, R., et al., Calculation of three-dimensional left ventricular strains from bi-planar tagged MR images, *Magn. Reson. Imaging* **2** (1992), 165.
20. Muthupillai, R., Lomas, D.J., Rosman, P.J., Greenleaf, J.F., Manduca, A., and Ehman, R.L., Magnetic resonance elastography by direct visualization of propagating acoustic strain waves, *Science* **269** (1995), 1854.

Functional Brain MRI

21. Bandettini, P.A., Jesmanowicz, A., et al., Processing strategies for time-course data sets in functional MRI of the human brain, *Magn. Reson. Med.* **30** (1993), 161–173.
22. Bandettini, P.A., Wong, E.C., Hinks, R.S., Tikofsky, R.S., and Hyde, J.S., Time course EPI of human brain function during task activation, *Magn. Reson. Med.* **25** (1992), 390–397.
23. Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., and Rosen, B.R., Functional mapping of the human visual cortex using magnetic resonance imaging, *Science* **254** (1991), 43–49.
24. Cohen, M.S., and Weisskoff, R.M., Ultra-fast imaging, *Magn. Reson. Imaging* **9** (1991), 1–37.
25. Hajnal, J.V., Myers, R., et al., Artifacts due to stimulus correlated motion in functional imaging of the brain, *Magn. Reson. Med.* **31** (1994), 283–291.
26. Kennan, R., Zhong, J., and Gore, J., Intravascular susceptibility contrast mechanisms in tissues, *Magn. Reson. Med.* **31** (1994), 9–21.
27. Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H.M., Brady, T.J., and Rosen, B.R., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation, *Proc. Natl. Acad. Sci. USA* **89** (1992), 5675–5679.

28. McCarthy, G., Blamire, A.M., Rothman, D.L., Gruetter, R., and Shulman, R.G., Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans, *Proc. Natl. Acad. Sci. USA* **90** (1993), 4952–4956.
29. Menon, R.S., Ogawa, S., Tank, D.W., and Ugurbil, K., 4 tesla gradient-recalled echo characteristics of photic stimulation induced signal changes in the human primary visual cortex, *Magn. Reson. Med.* **30** (1993), 380–386.
30. Moonen, C.T., Liu, G., van Gelderen, P., and Sobering, G., A fast gradient-recalled MRI technique with increased sensitivity to dynamic susceptibility effects, *Magn. Reson. Med.* **26** (1992), 184–189.
31. Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W., Brain magnetic resonance imaging with contrast dependent on blood oxygenation, *Proc. Natl. Acad. Sci. USA* **87** (1990), 9868–9872.
32. Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S-G., Merkle, H., and Ugurbil, K., Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging, *Proc. Natl. Acad. Sci. USA* **89** (1992), 5951–5955.
33. Stehling, M., Turner, R., and Mansfield, P., Echo-planar imaging: Magnetic resonance imaging in a fraction of a second, *Science* **254** (1991), 43–50.
34. Turner, R., Jezzard, P., Wen, H., Kwong, K.K., Le Bihan, D., Zeffiro, T., and Balaban, R.S., Functional mapping of the human visual cortex at 4 tesla and 1.5 tesla using deoxygenation contrast EPI, *Magn. Reson. Med.* **29** (1993), 277–279.

Multinuclear MRI

35. Albert, M.S., Cates, G.D., Driehuys, B., et al., Biological magnetic resonance imaging using laser-polarized ^{129}Xe , *Nature* **370** (1994), 199–201.
36. Bottomley, P.A., MR spectroscopy of the heart: The status and challenges, *Radiology* **91** (1994), 593–612.
37. Bottomley, P.A., Human in vivo NMR spectroscopy in diagnostic medicine: Clinical tool or research probe?, *Radiology* **170** (1989), 1–15.
38. Brown, T.R., Kincaid, B.M., and Ugurbil, K., NMR chemical shift imaging in three dimensions, *Proc. Natl. Acad. Sci. USA* **79** (1982), 3523–3526.
39. Duyn, J.H., and Moonen, C.T., Fast proton spectroscopic imaging of human brain using multiple spin-echoes, *Magn. Reson. Med.* **30** (1993), 409–414.
40. Maudsley, A.A., Hilal, S.K., Perman, W.H., and Simon, H.E., Spatially resolved high resolution spectroscopy by four-dimensional NMR, *J. Magn. Reson.* **51** (1983), 147–152.
41. Middleton, H., Black, R.D., Saam, B., et al., MR imaging with hyperpolarized ^3He gas, *Magn. Reson. Med.* **33** (1995), 271–275.
42. Posse, S., De Carli, C., and Le Bihan, D., Three-dimensional echo-planar spectroscopic imaging at short echo times in the human brain, *Radiology* **192** (1994), 733–738.
43. Schwarzschild, B., Inhaling hyperpolarized noble gas helps magnetic resonance imaging of lungs, *Physics Today* **48** (June 1995), 17–18.

Microscopic Imaging

44. Aguayo, J., Blackband, S., Schoeniger, J., Mattingly, M., and Hintermann, M., Nuclear magnetic resonance imaging of a single cell, *Nature* **322** (1986), 190–191.
45. Ahn, C.B., and Chu, W.C., Optimal imaging strategies for three-dimensional nuclear magnetic resonance microscopy, *J. Magn. Reson.* **94** (1991), 455–470.
46. Callaghan, P.T., *Principles of Nuclear Magnetic Resonance Microscopy*, Oxford University Press, New York, 1991.
47. Cho, Z.H., Ahn, C.B., Juh, S.C., and Lee, H.K., Nuclear magnetic resonance microscopy with 4 micrometer resolution: Theoretical study and experimental results, *Med. Phys.* **15** (1988), 815–824.

48. Chung, H., Wehrli, F.W., Williams, J.L., and Kugelmass, S.D., Relationship between NMR transverse relaxation, trabecular bone architecture and strength, *Proc. Natl. Acad. Sci. USA* **90** (1993), 10250–10254.
49. Cofer, G.P., Brown, J.M., and Johnson, G.A., In vivo magnetic resonance microscopy at 5 micrometers, *J. Magn. Reson.* **83** (1989), 608–616.
50. Early, T., Roemer, P., Mueller, O., Mogro-Campero, A., Turner, L., and Johnson, G., A high-temperature superconducting receiver for nuclear magnetic resonance microscopy, *Science* **259** (1993), 793–795.
51. Gewalt, S., Glover, G., Hedlund, L., Cofer, G., MacFall, J., and Johnson, G., MR microscopy of the rat lung using projection-reconstruction, *Magn. Reson. Med.* **29** (1993), 99–106.
52. Wu, Z., Chung, H., and Wehrli, F.W., A Bayesian approach to subvoxel tissue classification in NMR microscopic images of trabecular bone, *Magn. Reson. Med.* **31** (1994), 302–308.
53. Xiang, Q.S., and Henkelman, R.M., K-space description for the MR imaging of dynamic objects, *Magn. Reson. Med.* **29** (1993), 422–428.

Labels for Figure 4.1.

Year

$$\log\left(\frac{1}{\text{voxel volume in mm}^3}\right)$$

Chapter 5

Single Photon Emission Computed Tomography

5.1 Introduction

Single photon emission computed tomography (SPECT) is a medical imaging technique that is based on conventional nuclear medicine imaging and tomographic reconstruction methods. The images reflect functional information about patients similar to that obtained with positron emission tomography (PET). Both SPECT and PET (see Chapter 6) give information based on the spatial concentration of injected radiopharmaceuticals, in contrast to the other medical imaging modalities used for clinical diagnostic purposes.

Figure 5.1 shows the imaging process and the components of a typical nuclear medicine imaging system. A radioactive-labeled pharmaceutical (radiopharmaceutical) is administered to a patient. Depending on the biodistribution properties of the radiopharmaceutical, it is taken up by different organs and/or tissue types. Most radiopharmaceuticals used in nuclear medicine and SPECT are labeled with radionuclides that emit γ -ray photons. Typically, a scintillation camera system is used as the imaging device. The scintillation camera consists of a lead collimator that allows photons traveling in given directions to pass through a large-area scintillator (commonly NaI(Tl) crystal) that converts the energy of γ -ray photons to lower-energy photons which are in turn converted to electric signals by photomultiplier tubes (PMTs). The signals from an array of PMTs are processed by electronic circuitry to provide information about the position at which a photon interacts with the crystal. The scintillation camera provides a two-dimensional projection image of the three-dimensional radioactivity distribution or radiopharmaceutical uptake within the patient.

SPECT takes conventional two-dimensional nuclear medicine images acquired at different views around the patient and provides an estimate of the three-dimensional radioactivity distribution using methods of image reconstruction from multiple projections. SPECT differs from x-ray computed tomography (CT) in that the radiation source is within instead of outside the patient. The goal of SPECT is to determine accurately the three-dimensional radioactivity distribution resulting from the radiopharmaceutical uptake inside the patient (instead of the attenuation coefficient distribution from different tissues as obtained from x-ray CT). SPECT utilizes radiopharmaceuticals that are common in nuclear medicine clinics, rather than those that emit positrons with subsequent generation of two 511-keV annihilation photons as is the case with PET. SPECT requires instrumentation and image reconstruction methods that differ from those used in other medical imaging modalities.

The amount of radiopharmaceutical that can be administered is limited by the allowable dose of radiation to the patient. This requirement results in a limited number of photons that can be used for imaging. Also, the acceptance angle or geometric response of the collimator further limits the fraction of photons that are acceptable for the projection data. The collimator can be designed to allow detection of more photons, but increased detection efficiency usually can be achieved only with a concurrent loss of spatial resolution. A major goal of SPECT instrumentation development is to increase the detection efficiency while at the same time improving the spatial resolution of the imaging system, goals that are pursued by adding more detectors around the patient.

The SPECT imaging process imposes unique difficulties and challenges in image reconstruction. For example, before exiting the patient, many of the γ -ray photons experience photoelectric interactions that cause absorption of photons, and many thus experience Compton scattering, which changes the direction and energy of the original pho-

Figure here Figure 5.1. The imaging process and the components of a typical nuclear medicine imaging system.

Figure here Figure 5.2. Simple schematic showing the development over time of SPECT systems.

tons. When conventional reconstruction techniques (e.g., x-ray CT algorithms) are used in SPECT, the reconstructed images are severely affected by statistical noise fluctuations, poor spatial resolution, low contrast, and inaccurate quantitative information.

Following is a brief discussion of the basic principles of SPECT imaging and the current status of SPECT instrumentation and image reconstruction methods. Emphasis is placed on the physics, mathematics, and engineering aspects of SPECT, and future trends and potential areas of further investigation are discussed. In combination with new radiopharmaceuticals and clinical applications, these developments could ultimately improve patient care.

5.2 Physical and Instrumentation Factors That Affect SPECT Images

Attenuation from photoelectric absorption and Compton scattering is the major factor that affects the quantitative accuracy and quality of SPECT images. The degree of attenuation is determined by (1) the path length between the source and the edge of the attenuating material, and (2) the linear attenuation coefficient, which is a function of photon energy and the amount and types of materials contained in the attenuating medium. For example, the attenuation coefficient for the 140-keV photons (from the commonly used isotope Tc-99m) in water or soft tissue is 0.15 cm^{-1} . For the 140-keV photons, this attenuation coefficient gives rise to a half-valued-layer—the thickness of material that attenuates half of the incident photons—of 4.5 cm in water. The attenuation effect is further complicated by the fact that different regions of the body have different attenuation coefficients. In particular, the non-uniform attenuation distribution in the thorax is a major problem in cardiac SPECT image reconstructions.

Photons that have been scattered before reaching the radiation detector give erroneous spatial information about the origin of the radioactive source, and a significant fraction of the photons detected in SPECT have been scattered: typical ratios of scattered to unscattered photons are about 20–30% in brain SPECT studies and about 30–40% in cardiac and body SPECT studies for 140-keV photons. Overall, the resolution of contemporary SPECT systems is about 7 mm. As this measurement is about the same as the thickness of many organ subregions (e.g., in the brain, breast, kidney, or bones), the resulting images are significantly blurred. This relatively low resolution affects image quality as well as quantitative accuracy.

5.3 SPECT Instrumentation

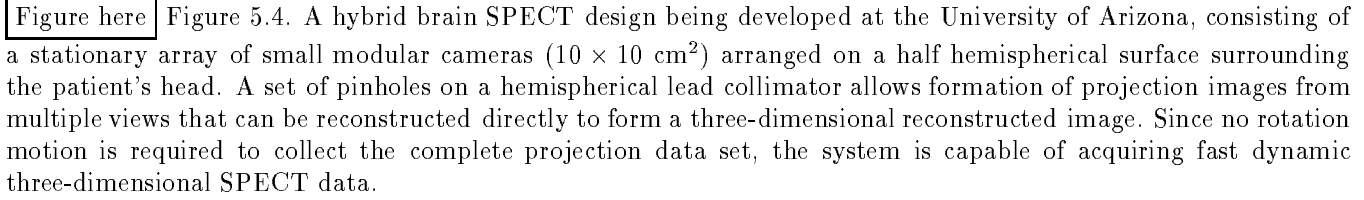
5.3.1 SPECT System Designs

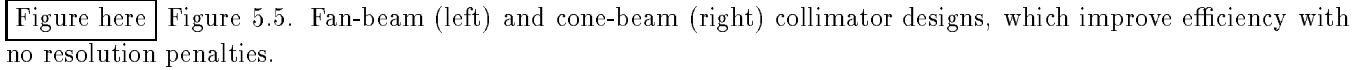
The development of SPECT systems can be summarized by the schematic diagram shown in Figure 5.2. The systems are designed to acquire projection data from multiple views around the patient. In general, SPECT instrumentation can be divided into three categories: those that use (1) arrays of multiple scintillation detectors, (2) one or more scintillation cameras, or (3) hybrid scintillation detectors combining the first two approaches.

Figure 5.2 (upper left) shows the first multiple-detector SPECT system design developed in the 1960s and consisting of four banks of linear arrays of detectors. Multidetector SPECT systems provide the capability of high counting rates but are restricted to single or a limited number of image slices. The SPECT system designs found in most commercial SPECT systems are based on a single scintillation camera or on multiple rotating scintillation cameras. The scintillation cameras allow acquisition of two-dimensional projection data that can be reconstructed to form multiple reconstructed image slices. Although the counting rate of the camera-based SPECT systems is generally lower than that of multiple detector-based SPECT systems, the data from camera-based SPECT systems are usually adequate for most clinical studies. Multiple-camera-based SPECT systems increase the number of cameras surrounding the patient to provide higher detection efficiency (Fig. 5.3). The three-camera design allows 360° sampling in periods as short as 5 s.

Figure here Figure 5.3. Different arrangements of multiple-camera-based SPECT systems, which increase the number of cameras surrounding the patient to provide higher detection efficiency.

A unique hybrid brain SPECT system design is being developed at the University of Arizona. As shown in Figure 5.4, the system consists of a stationary array of small modular cameras ($10 \times 10 \text{ cm}^2$) arranged on a half hemispherical surface surrounding the patient's head. A set of pinholes on a hemispherical lead collimator allows formation of projection images from multiple views that can be reconstructed directly to form a three-dimensional reconstructed image. Since no rotation motion is required to collect the complete projection data set, the system is capable of acquiring fast dynamic three-dimensional SPECT data.

 Figure 5.4. A hybrid brain SPECT design being developed at the University of Arizona, consisting of a stationary array of small modular cameras ($10 \times 10 \text{ cm}^2$) arranged on a half hemispherical surface surrounding the patient's head. A set of pinholes on a hemispherical lead collimator allows formation of projection images from multiple views that can be reconstructed directly to form a three-dimensional reconstructed image. Since no rotation motion is required to collect the complete projection data set, the system is capable of acquiring fast dynamic three-dimensional SPECT data.

 Figure 5.5. Fan-beam (left) and cone-beam (right) collimator designs, which improve efficiency with no resolution penalties.

5.3.2 Special Collimators

As shown in Figure 5.1, the most common collimator design used in nuclear medicine and SPECT consists of parallel holes. For the parallel-hole collimator design, the detection efficiency can be increased at the expense of a concurrent degradation of spatial resolution. This trade-off between detection efficiency and spatial resolution is a major factor in the development of SPECT systems and in the improvement of SPECT image quality. The goal of special collimator designs in SPECT is to decrease the severity of this trade-off as compared to the parallel-hole design.

Two collimator designs used in SPECT that improve efficiency with no resolution penalties are the fan-beam and cone-beam geometries (Fig. 5.5). The gain in detection efficiency is obtained by the increased solid angle of acceptance. Typically, the fan-beam and cone-beam collimators provide about 1.5 and 2 times the detection efficiency of a parallel-hole collimator with the same spatial resolution. However, the geometries of the converging-hole collimator designs require special reconstruction algorithms and data acquisition strategies.

5.3.3 New Radiation Detector Technologies

For over 3 decades the scintillation camera has been the most popular imaging device of nuclear medicine. During this time its performance has been thoroughly studied and optimized, and further improvement will likely be dependent on the development of solid-state detector arrays. Current scintillation cameras have about 3-mm intrinsic spatial resolution and 10% energy resolution at 140 keV, numbers that have remained essentially unchanged for over a decade. Since the spatial resolution of a commonly used general-purpose collimator is on the order of 7 to 15 mm, the 3-mm intrinsic resolution of the camera has been regarded as adequate. However, new detector and collimator strategies such as those embodied in the design shown in Figure 5.4 give encouragement that improved resolution can be achieved. Simulations from the research group at the University of Arizona have shown that a final tomographic resolution of around 2 mm is possible if detectors with very large areas and small pixels can be produced.

Semiconductor detectors or combinations of scintillators and solid state photon detectors offer good prospects for significant improvements in spatial and energy resolution. In the past, silicon and germanium detectors were investigated for use in conventional nuclear medicine imaging. However, their use was limited due to their lower stopping power, high material cost, and the requirement to operate at low temperature. However, new room-temperature semiconductors with high stopping power are now available, offering numerous options for new detector designs. For example, one exciting new material, cadmium zinc telluride, now yields a room-temperature energy resolution of about 3–4% at 140 keV. These new semiconductor detectors combined with optimized imaging geometries offer the prospect of major advances in the performance of SPECT imaging systems. The limit in useful resolution depends ultimately on the allowable patient dose of radionuclide and the solid angle effectively spanned by the detection system.

The sensitivity of SPECT is much less than that of PET because of the necessity of using lead channels or collimators, which limit the solid angle subtended by the tomograph. A general expression for sensitivity S is

$$S \propto \frac{A\epsilon^n\gamma}{4\pi r^2},$$

where A is the area of detector material seen by each point in the object being scanned, ϵ is the efficiency of the detector, $n = 1$ for SPECT and 2 for PET, γ is the attenuation factor (typically 0.2 to 0.6), and r is the radius of the tomograph. One can show that the ratio of sensitivity between PET and SPECT is about 150 divided by the resolution; thus, for a resolution of 7.5 mm, the ratio is 20.

Innovations suggested to improve the sensitivity of SPECT include not only methods of increasing A through collimator-detector arrangements, but also the use of methods for tracking the trajectory of each photon. The latter include Compton scattering angle determinations from energy discrimination.

5.4 SPECT Image Reconstruction

5.4.1 The SPECT Reconstruction Problem

In SPECT, the goal of image reconstruction is to determine accurately the three-dimensional distribution of administered radiopharmaceutical in the patient. Assume an ideal situation in which the emission photons do not experience attenuation and scatter in the patient and the collimator-detector has perfect spatial resolution without blurring effects in the measured data. A naive approach to the SPECT reconstruction is to consider the projections $p(t, \theta)$ of radiopharmaceutical activity as the simple Radon transform, which for a two-dimensional distribution $f(x, y)$ is

$$p(t, \theta) = c \int_{-\infty}^{+\infty} f(x, y) ds, \quad (5.1)$$

where t is the position on the projection array, θ is a particular projection angle, and c is the gain factor that transforms radioactivity concentration to detected signals. The SPECT reconstruction problem is different from the classical problem of image reconstruction from projections because each photon is attenuated and scattered by the material between its source and the detector.

When attenuation is taken into consideration, the two-dimensional attenuated Radon transform can be written as

$$p(t, \theta) = c \int_{-\infty}^{+\infty} f(x, y) \exp\left[\int_{(x,y)}^{+\infty} a(u, v) ds'\right] ds, \quad (5.2)$$

where $a(u, v)$ is the two-dimensional attenuation coefficient distribution and $\int_{(x,y)}^{+\infty} a(u, v) ds'$ is the attenuation factor for photons that originate from (x, y) , travel along the direction perpendicular to the detector array, and are detected by the collimator-detector. See also section 14.1.2.

However, in realistic situations, the projection data in SPECT are often measured in two dimensions and are affected by the three-dimensional effects of attenuation and scatter of photons in the patient's body and the three-dimensional spatial resolution response of the collimator-detector. When these three-dimensional effects are taken into consideration, the measured projection data are given by a more complicated attenuated Radon transform,

$$p(\mathbf{t}, \theta) = c \int_{\Omega} h(s, \omega; \mathbf{r}) \int_{-\infty}^{+\infty} f(\mathbf{r}) \exp\left[-\int_{\mathbf{r}}^{+\infty} a(\mathbf{u}) ds'\right] ds d\omega, \quad (5.3)$$

where $\mathbf{t} = (x, y)$ is a point on the two-dimensional projection image $p(\mathbf{t}, \theta)$ at viewing angle θ , and $f(\mathbf{r})$ and $a(\mathbf{u})$ are the three-dimensional radioactivity and attenuation distributions in the patient, respectively. At each point on the projection image, unattenuated and scattered photons that fall within the field of view of the collimator-detector are detected. The inner integral of equation 5.3 considers attenuation of these photons through the patient with a three-dimensional attenuation distribution $a(\mathbf{u})$. The combined geometric collimator-detector and scatter response are represented by a three-dimensional response function, $h(s, \omega; \mathbf{r})$, which is a function of the position \mathbf{r} at which the photon originates in the patient and its distance from the collimator-detector. The function is non-zero within a solid angle defined by Ω .

5.4.2 SPECT Image Reconstruction Methods

The SPECT reconstruction problem without attenuation expressed in equation 5.1 estimates $f(x, y)$ by using conventional techniques for reconstructing images from projections. The most commonly used techniques are based on the Fourier projection theorem. These techniques apply a ramp-shaped filter to the Fourier transform of the projections (equation 5.1) and then backproject the inverse transform. The methods are generally known as filtered backprojection of filtered projection or, equivalently, convolution methods. The shape of the filter depends on the amount of noise suppression desired. (More detailed descriptions of these image reconstruction techniques are found in Chapter 14.) When used in routine clinical SPECT image reconstruction, these reconstructions ignore the effects of attenuation, scatter, and the collimator-detector response. Consequently, SPECT images obtained with these reconstruction techniques are quantitatively inaccurate. More importantly, these images have artifacts and distortion, resulting in poor image quality. However, analytical solutions to the inverse Radon transform expressed in equations 5.2 and 5.3 have not been found for the general case of non-uniform attenuation. Hence, special reconstruction methods are needed to estimate accurately the true three-dimensional radiopharmaceutical distribution in the patient, and research in this area has been a major focus in the development of SPECT in recent years. In particular, iterative reconstruction algorithms have been used to provide estimates of the solutions of the SPECT reconstruction problem expressed in equations 5.2 and 5.3.

In general, quantitative SPECT image reconstruction methods consist of two major components: (1) algorithms for image reconstruction from projections, and (2) techniques that compensate for image-degrading effects such as

attenuation, scatter, and collimator-detector response. Often, the image reconstruction algorithm is inseparable from the compensation technique, resulting in SPECT reconstruction methods that are not found in other tomographic imaging modalities.

Iterative Reconstruction Algorithms A typical iterative reconstruction algorithm starts with an initial estimate of the object source distribution. Projection data are generated from the initial estimate using a projection matrix that models the imaging process. The calculated projection data are compared with the measured data at the same projection angles and their differences are determined. With application of a given statistical criterion, the differences are used to update the initial image estimate. A new set of calculated projection data is regenerated from the new image estimate and compared with the measured projection data. The procedure is repeated until the differences between the calculated and measured projection data are smaller than a preselected small value. Statistical criteria that have been used in formulating iterative reconstruction algorithms include the minimum mean squares error (MMSE), weighted least squares (WLS), maximum entropy (ME), maximum likelihood (ML), and maximum a posteriori (MAP) approaches. Iterative algorithms that have been used in estimating the reconstructed images include the Gauss-Seidel (GS), conjugate gradient (CG), and expectation maximization (EM) algorithms (see section 14.1.2). Iterative reconstruction methods differ in their assumptions about the statistical properties of the measured data, convergence rates, and noise characteristics. These properties are important considerations for assessing the applicability of particular reconstruction methods in SPECT.

Interest in the applications of iterative reconstruction algorithms was initially propelled by the need to compensate for non-uniform attenuation in the imaging of heart muscle with radiopharmaceuticals, which illuminate relative perfusion. More recently, iterative reconstruction algorithms have been applied to compensate for the spatially variant system response function (collimator-detector response) and scatter. These compensations are achieved by modeling the imaging process in the projection and backprojection operations during the iterative steps. A major drawback of the iterative techniques is the long processing time involved. However, rapid advances in computer technology and in software implementation of these algorithms have significantly reduced the computation times. For example, the reconstruction time for 30 iterations of the iterative ML-EM algorithm with non-uniform attenuation compensation is 4 minutes for 32 64×64 image slices using a state-of-the-art DEC Alpha 3000 workstation. Continued development in computer hardware, algorithms, and codes will bring quantitative SPECT reconstruction methods to clinical use in the near future.

Compensation Methods Methods of compensating for attenuation can be grouped into two categories: (1) methods that assume that the attenuation coefficient is uniform over the body region, and (2) methods for non-uniform distributions of attenuation coefficients. The assumption of uniform attenuation can be applied to SPECT imaging of the head and of the abdomen region. If uniform attenuation is assumed, i.e., $a(x, y) = \text{constant}$, analytical solution of the inverse attenuated Radon transform given in equation 5.2 can be derived, and the radioactivity distribution can be estimated accurately.

In cardiac SPECT studies, iterative reconstruction methods have been used to compensate for the non-uniform attenuation distribution that arises from the heterogeneity of the chest region. Transmission CT methods using an external radiation source in the SPECT instrument can map the attenuation coefficient distribution in a patient. The attenuation coefficient distribution is then incorporated into the iterative reconstruction algorithms to achieve accurate compensation for the attenuation effects. The resultant SPECT images show improved quantitative accuracy and fewer image artifacts and distortions.

Methods of compensating for scatter can be grouped into two general approaches. In the first approach, scattered photons are considered to carry no useful imaging information. The compensation techniques involve estimating the scatter component and subtracting it from the measured data or from the reconstructed images to obtain scatter-free reconstructed images. Various techniques have been developed, including those using two or more energy windows during acquisition. The compensation methods based on subtraction of scattered photons provide SPECT images with increased contrast and improved quantitation. However, the methods tend to increase noise in the subtracted images.

In the second approach, the scattered photons are utilized in estimating the true radioactivity distribution. When the scatter component is not subtracted, the compensated images are less noisy than those obtained with the first approach. In one method, an average scatter response function is combined with the geometric response of the collimator-detector to form the total response of the imaging system. The total response function is then used to generate a restoration filter that approximately compensates for the geometric and scatter response. Another class of scatter compensation methods determines the exact scatter response function and incorporates it into iterative reconstruction algorithms. Since the exact scatter response function depends on the physiology of a particular patient as well as on the characteristics of the imaging system, it is nonstationary and asymmetric in shape. The determination and implementation of these scatter compensation methods require extensive computations. However,

efforts are being made to parameterize the scatter response function and to optimize its implementation in the iterative reconstruction methods to achieve a substantial reduction in processing time.

By assuming an average and stationary geometric collimator-detector response function, restoration filters can be used to provide partial and approximate compensation for the effects of a collimator-detector. Examples are the Metz and Wiener filters. An analytical method has also been developed to compensate for the spatially variant geometric collimator-detector response under the assumption that the response function has a special mathematical form. In addition, the exact collimator-detector response function can be incorporated into iterative reconstruction methods for accurate compensation of scattering.

Three-Dimensional Reconstruction Methods for Special Collimator Designs As discussed in the previous section, fan-beam and cone-beam collimators have been used in SPECT to improve the trade-off between detection efficiency and spatial resolution as compared to the conventional parallel-hole design. Due to the special geometries of such converging-hole collimators (others include astigmatic and variable focal collimators), special three-dimensional reconstruction methods and acquisition strategies are required.

For example, the typical cone-beam collimator shown in Figure 5.5 provides an example of a fully three-dimensional SPECT system configuration where the paths of the detected photons lie on lines passing through the collimator focal point and cannot be grouped into two-dimensional transverse planes. The set of cone-beam projections taken from a 360° circular motion of the detector does not provide all the necessary samples for a complete three-dimensional reconstruction. As a result, reconstruction from this geometry constitutes a three-dimensional “limited-angle reconstruction” problem, and the reconstructed images have artifacts and distortions as a result. Special three-dimensional reconstruction methods and data acquisition strategies are being developed to overcome this problem. One approach is to use interpolation techniques to fill part of the missing region and thus reduce artifacts due to missing data. Another approach is to gather a complete set of cone-beam projections using specially designed orbits. To ensure a complete projection data set, the path in space taken by the focal point of the cone-beam collimator must intersect every plane crossing the reconstructed image space by having rotation orbits that are not constrained to lie in a plane. Optimizing detector motion, while satisfying the orbit requirements and avoiding data truncation, is an area of current research in SPECT.

5.5 Research Opportunities

Promising research and development directions related to innovative detector technologies include the following:

- Investigation of new room-temperature semiconductor materials, combination of scintillators and solid-state detectors, and readout electronics for use in single photon imaging systems; and
- Development of digital scintillation cameras with high counting rates and improved imaging capabilities, including improved stability, uniformity, energy resolution, and intrinsic resolution.

In the area of better SPECT system configurations, research is recommended for the following:

- Development of new SPECT system configurations for improved detection efficiency, spatial resolution, and ease of operation; and
- Development of new collimator designs for improved trade-off between spatial resolution and detection efficiency.

The most promising research directions related to image reconstruction and mathematical approaches include the following:

- Development of analytical solutions for the SPECT reconstruction problem that include the effects of non-uniform attenuation, spatially varying scatter, and collimator-detector response;
- Development of fast and stable iterative reconstruction methods that incorporate non-uniform attenuation, spatially varying scatter, and collimator-detector response for accurate SPECT image reconstruction; and
- Development of new three-dimensional reconstruction methods and data acquisition strategies for converging-beam SPECT.

5.6 Suggested Reading

1. Barrett, H.H., Perspectives on SPECT, *Proc. SPIE* **671** (1986), 178–183.

2. Budinger, T.F., Gullberg, G.T., and Huesman, R.H., Emission computed tomography, in *Image Reconstruction from Projections: Implementation and Applications*, G.T. Herman, ed., Springer-Verlag, New York, 1979, 147–246.
3. Gullberg, G.T., Zeng G.L., Datz, F.L., Christian, P.E., Tung, C.-H., and Morgan, H.T., Review of convergent beam tomography in single photon emission computed tomography, *Phys. Med. Biol.* **37** (1992), 507–534.
4. Jaszczak, R.J., and Tsui, B.M.W., Single photon emission computed tomography, in *Principles of Nuclear Medicine*, H.N. Wagner, Jr., ed., 2nd Edition, W.B. Saunders Company, Philadelphia, 1995.
5. Shepp, L.A., and Vardi, Y., Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging* **MI-1** (1982), 113–122.
6. Tretiak, O.J., and Metz, C.E., The exponential Radon transform, *SIAM J. Appl. Math.* **9** (1980), 341–354.
7. Tsui, B.M.W., Frey, E.C., Zhao, X-D., Lalush, D.S., Johnston, R.E., and McCartney, W.H., The importance and implementation of accurate 3D compensation methods for quantitative SPECT, *Phys. Med. Biol.* **39** (1994), 509–530.
8. Tsui, B.M.W., Zhao, X-D., Frey, E.C., and McCartney, W.H., Quantitative single-photon emission computed tomography: Basics and clinical considerations, *Semin. Nucl. Med.* **XXIV** (1994), 38–65.

Chapter 6

Positron Emission Tomography

6.1 Introduction

6.1.1 History

The history of positron emission tomography (PET) can be traced to the early 1950s, when the medical imaging possibilities of a particular class of radioactive substances were first realized. It was recognized then that the high-energy photons produced by the annihilation of the positron-emitting isotopes could be used to describe, in three dimensions, the physiological distribution of “tagged” chemical compounds. After 2 decades of moderate technological developments by a few research centers, widespread interest and broadly based clinical research activity began in earnest following the development of sophisticated reconstruction algorithms and improvements in detector technology. By the mid-1980s, PET had become a tool for medical diagnosis and for dynamic studies of human metabolism. Table 6.1 indicates the significant advances in PET technology over the last 20 years.

6.1.2 Applications

Research with PET has added immeasurably to our current understanding of flow, oxygen utilization, and the metabolic changes that accompany disease and that change during brain stimulation and cognitive activation. Clinical uses include studies of Alzheimer’s disease, Parkinson’s disease, epilepsy, and coronary artery disease affecting heart muscle metabolism and flow. Even more promising with regard to widespread clinical utilization of PET are recent developments showing that PET can be used effectively to locate tumors and metastatic disease in the brain, breast, lung, lower gastrointestinal tract, and other sites. Early evidence also indicates that quantitative studies of tumor metabolism with PET can be used for non-invasive staging of the disease. Compared to other cross-sectional imaging techniques like MRI and CT, PET is distinguished by its immense sensitivity—its ability to quantitatively determine and display tracer concentrations in the nanomolar range.

6.1.3 Principle of Operation

PET imaging begins with the injection of a metabolically active tracer—a biological molecule that carries with it a positron-emitting isotope (for example, ^{11}C , ^{13}N , ^{15}O , or ^{18}F). Over a few minutes, the isotope accumulates in an area of the body for which the molecule has an affinity. For example, glucose labeled with ^{11}C , or a glucose analog labeled with ^{18}F , accumulates in the brain or in tumors, where glucose is used as the primary source of energy. The radioactive nuclei then decay by positron emission. The ejected positron combines with an electron almost instantaneously, and these two particles undergo the process of annihilation. The energy associated with the masses of the positron and electron is divided equally between two photons that fly away from one another at a 180° angle. Each photon has an energy of 511 keV. These high-energy γ -rays emerge from the body in opposite directions, to be detected by an array of detectors that surround the patient (Fig. 6.1). When two photons are recorded simultaneously

Table 6.1 Improvements in PET Technology Since 1975

Parameter	Typical Values, 1975	Typical Values, 1995
Spatial resolution	14 mm	4 mm
Number of detectors	64	19,000
Data per study	4 kilobytes	4 megabytes

Figure here Figure 6.1. Arrangement of PET detectors in a ring camera surrounding the patient port.

Figure here Figure 6.2. Block detector principle of operation. The position histogram of the ratios of light intensities x and y allows the identification of the crystal of interaction.

by a pair of detectors, the annihilation event that gave rise to them must have occurred somewhere along the line connecting the detectors. Of course, if one of the photons is scattered, then the line of coincidence will be incorrect. After 100,000 or more annihilation events are detected, the distribution of the positron-emitting tracer is calculated by tomographic reconstruction procedures from the recorded projection data:

$$P = e^{-\int a(s) ds} \int f(s) ds, \quad (6.1)$$

where P is the projection data, $a(s)$ is the linear attenuation coefficient for 511-keV γ -rays, and $f(s)$ is the isotope distribution. Section 14.1.2 provides another look at the origin of this equation.

The exponential factor in equation 6.1 is required as a means to take into account the attenuation of the two γ -rays inside the patient's body. The effect of scattered radiation, although significant in certain imaging situations, is ignored in equation 6.1.

PET reconstructs a two-dimensional image from the one-dimensional projections seen at different angles. Three-dimensional reconstructions can also be done using two-dimensional projections from multiple angles. After appropriate corrections for both attenuation and scatter of the γ -rays within the body, the resulting images give a quantitative measure of the isotope distribution. PET images, therefore, can be used both to qualitatively assess the site of unusual tracer accumulations (e.g., in tumors) and to quantitatively measure the tracer uptake for a more in-depth diagnosis or staging of the disease.

6.2 Current Status of PET Technology

6.2.1 γ -Ray Detectors

Efficient detection of the annihilation photons from positron emitters is usually provided by the combination of a crystal, which converts the high-energy photons to visible-light photons, and a photomultiplier tube that produces an amplified electrical current pulse proportional to the number of light photons interacting with the photocathode. The fact that the imaging system's sensitivity in measuring coincident events is proportional to the square of the detector efficiency leads to a very important requirement that the detector be nearly 100% efficient. Thus, other detector systems such as plastic scintillators or gas-filled wire chambers, with typical individual efficiencies of 20% or less, would result in a coincident efficiency of only 4% or less.

Most modern PET scanners are multilayered with up to 63 levels or transaxial layers to be reconstructed. In the most common current application, tungsten or lead septa between the layers of detector elements reject events between different rings of detectors and thus decrease the sensitivity of the instrument, but also help to reject events in which one or both of the 511-keV photons suffer a Compton scatter within the patient.

The "individually coupled" detector design with one phototube per crystal is capable of very high data throughput, since the design is parallel (all photomultiplier tubes and scintillator crystals operate independently). The disadvantage of this type of design is the requirement for many expensive photomultiplier tubes. Additionally, connecting round photomultiplier tubes of sufficiently small diameter to rectangular scintillation crystals in order to form a solid ring leads to problems in packaging.

The contemporary method of packaging many scintillators for 511 keV around the patient is to use what is called a block detector design. The arrangement of scintillators and phototubes is shown in Figure 6.2. A block detector couples several photomultiplier tubes to a bank of scintillator crystals and uses a coding scheme to determine the crystal of interaction. Most block detector coding schemes use the ratio of light output between a number of photomultipliers to determine the crystal of interaction. In the example shown in Figure 6.2, four photomultiplier tubes are coupled to a block of bismuth germanate (BGO) that has been partially sawed through to form 64 "individual" crystals. The depth of the cuts is critical; that is, deep cuts tend to focus the scintillation light onto the face of a single photomultiplier tube, whereas shallower cuts tend to spread the light over all four photomultiplier tubes.

BGO, the most commonly employed scintillator in modern PET systems, has a light output of approximately 3000 photons per 511 keV. For statistical reasons this limits the number of individual crystals that can be decoded without ambiguity within a detector block that spreads the light into the area of a number of crystals. Improving

Figure here Figure 6.3. Loss in radial spatial resolution for off-axis events due to oblique incidence angle for γ -rays on the detector elements. (Courtesy Thomas F. Budinger, Lawrence Berkeley National Laboratory.)

the spatial resolution of block detectors without increasing the number of expensive photomultiplier tubes thus requires the use of scintillators with higher light output. A candidate that is currently under investigation is lutetium oxyorthosilicate (LSO) with a light output approximately 5 times higher than that of BGO. Another possible solution is to use higher-energy-photons (> 1 MeV) for detector calibration while allowing a certain statistical uncertainty when actually detecting 511-keV γ -rays. The resulting spatial detector response function will be wider than the width of the individual crystals, but due to better sampling the resulting image spatial resolution should still be improved.

Another new and promising aspect in PET detector technology is the replacement of photomultipliers by less expensive avalanche photodiodes (APDs). Although it has been shown that APDs allow achievement of the necessary time and pulse height resolution, the development of large and tightly packaged avalanche photodiode arrays still represents a technological challenge.

6.2.2 Limitations of the Spatial Resolution

Aside from the obvious effect of the size of the detector elements on the spatial resolution of PET systems, there are two fundamental effects limiting the resolution of the reconstructed images:

1. The angle between the paths of the annihilation photons can deviate from 180° as a result of some residual kinetic motion (Fermi motion) at the time of annihilation. The effect on resolution of this deviation increases as the detector ring diameter increases, so that eventually this factor can have a significant effect. For instance, with a diameter of 80 cm the effect on resolution is 1.7 mm.
2. The distance the positron travels after being emitted from the nucleus and before annihilation occurs is a source of deterioration in spatial resolution. This distance depends on the particular nuclide. For example, the range of blurring for ^{18}F , the isotope used for many of the current PET studies, is quite small (about 0.5 mm) compared with that of other isotopes.

When both effects are combined and infinite resolution of the detectors is assumed, the fundamental limit for the spatial resolution in systems whose apertures allow investigation of the whole human body is about 2 mm for ^{18}F and 5 mm for ^{82}Rb . Commercial systems with high-resolution block detectors today achieve a spatial resolution of approximately 4 mm when using ^{18}F .

For events outside the center or axis of the tomograph, the spatial resolution will be less by a significant amount. The path of the photons from an “off-center” annihilation typically traverses more than one detector crystal, as shown in Figure 6.3. This results in an elongation of the resolution spread function along the radius of the transaxial plane. The loss of resolution is dependent on the crystal density and the diameter of the tomograph detector ring. For an 80-cm-diameter system, the resolution can deteriorate by about 20% from the axis to 10 cm off axis. Any improvement in the resolution of the detectors beyond the currently achieved 4 mm, therefore, is meaningful only if it is combined with a method to determine the depth of interaction for each event within the individual detector crystals.

6.2.3 System Electronics

The electronics must be able to determine coincident events with about 10-ns resolution for each crystal-crystal combination (i.e., line of response). The timing requirement is set jointly by the time of flight across the detector ring (3 ns) and the crystal-to-crystal time resolution (typically 4 ns). The most stringent requirement, however, is the vast number of lines of response in which coincidences must be determined (over 1.5 million in a 24-layer camera with septa in place, and 18 million with the septa removed in order to also evaluate events between different detector rings).

It is obviously impractical to have individual coincidence circuitry for each line of response, and so tomograph builders use parallel organization to solve this problem. A typical method is to use a high-speed clock to mark the arrival time of each 511-keV photon based on this time marker. This search can be done extremely quickly by having multiple dedicated processors working in parallel.

A modern commercial PET system employs a total of approximately 18,000 detector crystals grouped into about 300 block detectors. Three or four rings of block detectors allow an axial field of view of 15 cm to be measured without repositioning the patient. In a typical high-resolution scanner the individual detector crystals form 32 rings of, for example, 576 crystals in each ring. To improve the reliability and reduce the cost of the system electronics,

CMOS application-specific integrated circuits (ASICs) have been developed for the front-end analog electronics, the position/energy logic of each detector block, and the parallel coincidence detection processors. Only by using such highly integrated electronics has it become possible to develop clinical PET scanners that have all the electronics circuitry inside the scanner gantry and a total power dissipation of only about 1 kW.

If two annihilation events occur within the time resolution of the tomograph (e.g., 12 ns), then “random” coincidence events add erroneous background activity to the data; this can be significant at high count rates. These events can be corrected for on a line-of-response by line-of-response basis. The random coincidence event rate of each crystal pair is measured by observing the rate of events in a delayed coincidence timing window. The random event rate for a particular line of response corresponding to a crystal pair is proportional to the event rates of each crystal and the coincidence window width. As the activity in the subject increases, the event rate in each detector increases. Thus, the random event rate will increase as the square of the activity.

The detected coincidence events need to be stored during acquisition at rates of up to 3 million per second. Rather than storing the events in chronological sequence (list mode), most modern systems provide real-time histogramming circuitry that allows data accumulation in projective views of the subject as required for the subsequent reconstruction. For each incoming coincidence pair, the correct address in projection space is determined in real time and one event is added to the corresponding memory location. This circuitry allows correction of random events in real time by subtracting the delayed coincidence pairs or “randoms” from the appropriate projection space data address. With real-time histogramming the data are available for immediate reconstruction after the end of the data acquisition without having to go through a time-consuming off-line sorting procedure.

6.2.4 Data Correction and Reconstruction Algorithms

Before reconstruction, each projection address or line of response receives three corrections: for crystal efficiency, attenuation, and dead time. The efficiency for each line of response is computed by dividing the observed count rate for that line of response by the average count rate for lines of response with similar geometry (i.e., length). The efficiency typically is measured using a radioactive ring or rotating rod source without the patient in place. Once the patient is in position, a transmission scan is taken with the same source and the attenuation factor for each line of response is computed by dividing its transmission count rate by its efficiency or “blank” count rate. If rotating rod sources are used instead of a ring source to acquire the blank and transmission scan, it is possible to geometrically distinguish lines of response that constitute the rod sources from other coincidence events that are due to scatter or background activity in the patient. In clinical practice this capability allows the transmission scan to be taken after the isotope injection and after the emission scan, which represents significant time savings, since the take-up time for the radiopharmaceutical may be as long as 1 hour. After transmission and emission data have been collected, the random-corrected emission rate is divided by the attenuation factor and efficiency for each line of response. The resulting corrected projection data are reconstructed, usually with the filtered backprojection algorithm. This is the same algorithm used in x-ray computed tomography. The projection data are formatted into parallel or fan-beam data sets for each angle. These are modified by a high-pass filter and backprojected.

Most of the conventional reconstruction techniques for implementing the inverse Radon transform are derived from the Fourier projection theorem. These methods involve application of a ramp-like frequency filter to the Fourier coefficients of each projection. After inverse transformation, the projections are backprojected and superposed to form the reconstructed image. The process of PET reconstruction is linear and is shown by operators successively operating on the projections P :

$$f = \sum_{\theta} \text{BP} F^{-1}[\text{R}F(f)],$$

where f is the image, F is the Fourier transform, R is the ramp-shaped high-pass filter, F^{-1} is the inverse Fourier transform, and BP is the backprojection operator.

An alternative class of reconstruction algorithms involve iterative solutions to the classic inverse problem

$$P = Af,$$

where P is the projection matrix, f is the matrix of the true image data being sought, and A is the projection operation. The inverse, $f = A^{-1}P$, is computed by iteratively estimating the data f' and modifying the estimate by comparison of the calculated projection set P' to the true observed projections P . The expectation-maximization algorithm solves the inverse problem by updating each pixel value f_i in accord with

$$f_i^{k+1} = \sum_j p_j \frac{f_i^k a_{ij}}{\sum_{i'} f_i^k a_{i'j}},$$

where p is the measured projection, a_{ij} is the probability that a source at pixel i will be detected in projection detector j , and k is the iteration (cf. section 14.1.2).

Figure here Figure 6.4. If the septa are removed to allow the acquisition of a larger number of lines of response in three-dimensional mode, the sensitivity increases by up to a factor of five. (Reprinted with permission from Cherry, S.R., Dahlbom, M., and Hoffman, E.J., 3D PET using a conventional multislice tomograph without septa, *J. Comput. Assisted Tomography* **15** (1991), 655–668. Copyright 1991 by Lippincott Raven Publishers.)

6.3 Three-Dimensional Acquisition and Reconstruction

6.3.1 Principle of Three-Dimensional Acquisition

The design of a multislice tomograph with interplane septa restricts the percentage of annihilation photon pairs detected from a central line source to less than 0.5% in the direct planes (i.e., acquiring coincidence data only within the same crystal ring) and an additional 0.75% in the cross planes (i.e., using the detector pairs of adjacent rings to form interplane slices). The most efficient use of the available photon flux can be realized by removing the interplane septa, which usually shield the detectors from the more oblique lines of response, and reprogramming the data acquisition hardware to accept all possible coincidence events (Fig. 6.4). This three-dimensional data acquisition technique requires the implementation of a three-dimensional algorithm for image reconstruction and the solution of problems associated with increased scatter and increased random coincidences. The increase in sensitivity, however, is very large, up to a factor of four to five owing primarily to the increased number of lines of response. The removal of the shadowing effect of the septa on the detectors further improves sensitivity. As a result, the dose of the radioisotope injected into the patient can be reduced, or the data acquisition time can be shortened to improve the patient's comfort without loss of image quality.

6.3.2 Three-Dimensional Reconstruction

PET systems allowing fully three-dimensional acquisition have a geometry equivalent to that of a truncated cylindrical detector. In a continuous model of such scanners, the integral of the unknown tracer distribution is measured along all straight lines that intersect the truncated cylinder in two points. These line integrals can be arranged in groups of parallel lines, to form a set of two-dimensional parallel projections of the object. This three-dimensional data set has three key properties:

1. Two-dimensional projections of the object are not measured for all orientations in space. The reconstruction, therefore, is a limited-solid-angle problem. This problem, however, has a unique and stable solution and therefore differs in nature from the two-dimensional limited-angle problem, which is known to be severely ill-posed (see section 14.4).
2. Mathematically, the data are highly redundant because a subset of the data, corresponding to the straight projections equivalent to the data measured in two-dimensional mode, is in principle sufficient to reconstruct the image. However, the improvement in sensitivity allowed by three-dimensional acquisition would be useless if only a subset of the data were incorporated into the reconstruction. Three-dimensional algorithms, therefore, all attempt to use the whole set of redundant data.
3. Most two-dimensional parallel projections are truncated: for all orientations where the lines of response are not perpendicular to the scanner axis, the integral of the tracer distribution is not measured for all lines crossing the field of view. All algorithms available today solve the truncation problem by exploiting, implicitly or explicitly, the possibility (property 2) of first reconstructing an estimate of the image based on a subset of the data.

The reprojection algorithm of Kinahan and Rogers, a three-dimensional filtered-backprojection algorithm that has been implemented by several groups, has been tested intensively since 1989. This algorithm uses the lower-statistics image obtained by reconstructing only the straight projection data in order to obtain the missing line integrals for the oblique data through forward projection. Tests have demonstrated an image quality equal to the quality obtained from a two-dimensional reconstruction, and they also demonstrate the ability of the reprojection algorithm to take full advantage of the increased sensitivity allowed for by the three-dimensional acquisition, i.e., the ability to translate that increased sensitivity into an improved image signal-to-noise ratio. The reprojection algorithm is an operational, fully tested solution to the three-dimensional image reconstruction problem. The reconstruction time required by the reprojection algorithm and the data storage requirements are obstacles to a more widespread application of three-dimensional acquisitions, in particular in the case of dynamic multiframe studies or whole-body acquisitions.

Some relief is possible by making use of the fact that the projection data from ring tomographs are naturally oversampled in the angular direction. This allows for a reduction in data without loss of spatial resolution by adding up projection data for adjacent angles, both for the axial and transaxial angles. Using this method of data reduction, an implementation of the reprojection algorithm on eight Intel i860 processors working in parallel on

different projections of the data set requires about 12 minutes in order to reconstruct a three-dimensional image set composed of 63 axial planes on a scanner with 32 rings of detector elements.

Other algorithms have been proposed to reduce the computational burden of three-dimensional reconstruction. The direct Fourier reconstruction method can be generalized to three dimensions, using the three-dimensional central section theorem: The two-dimensional Fourier transform of a two-dimensional parallel projection of an image is equal to a central planar section through the three-dimensional Fourier transform of that image. An interesting property of the data sampling provided by cylindrical scanners is that the interpolation from the polar grid to the three-dimensional Cartesian grid is two-dimensional (as in the two-dimensional case) rather than three-dimensional as could be expected a priori. The implementation of the direct Fourier reconstruction has been investigated. Good-quality reconstructions have been achieved with a gain in speed not exceeding a factor of two, but the algorithm is rather complex, and further work is needed to determine whether sufficiently accurate interpolation schemes could be used without losing the potential speed advantage of Fourier methods. The use of interpolation based on linograms has been proposed.

The FAVOR algorithm is a three-dimensional filtered backprojection algorithm that is based, as is the reprojection algorithm, on an exact analytical inversion. Although it has not been tested as intensively as the reprojection method, this algorithm appears to be as accurate, and to yield almost the same image signal-to-noise ratio for scanners with moderate axial apertures (smaller than 30°). The algorithm does not require forward projection and is about 30% faster than the reprojection algorithm.

6.3.3 Scatter Correction in Three Dimensions

Because of the removal of the interplane septa, the fraction of scattered events in three-dimensional mode is as high as 40% of all events detected, as compared with only about 12–14% in two-dimensional mode with the septa in place. This necessitates improvements of the scatter correction routines that in two dimensions usually just rely on simple deconvolution methods. One way is to employ the admittedly limited energy resolution of PET detectors by simultaneously acquiring data with a lower energy window representing mostly scattered events and a higher energy window for the unscattered events. Various algorithms have been developed to estimate the contamination of the data in the lower window with unscattered events and vice versa, as well as to overcome statistical noise problems associated with a reduced number of events in each window. The dual-energy based scatter correction techniques have been demonstrated to allow a reduction of the scatter-induced error in three dimensions to less than 10%. An inherent advantage of the dual-energy methods is their ability to account for scatter originating from “invisible” sources outside the scanner’s field of view. It is questionable, however, if this quantitatively small effect warrants the intrinsic increase in noise and added computational burden.

Another class of scatter correction techniques employs model-based estimates of the scatter. The uncorrected images serve as an estimate of the source distribution, while the transmission images represent the scattering medium. An estimate of the scatter contamination in the projection data then can be computed using the Klein-Nishina formula (an equation that describes Compton scatter) in a calculation of the scatter distribution. Tests of these methods have also been shown to correct the scatter related error in three-dimensional images to within about 5%. The forward calculation based techniques for scatter correction have not been attempted earlier, because of their computational complexity. With the computing power available now for three-dimensional reconstruction, however, such calculations can be performed with acceptable computation times, if appropriate sampling techniques are used.

6.3.4 Attenuation Correction in Three Dimensions

Transmission scans in three dimensions suffer from the fact that, because of the increase in count rate associated with the removal of the septa, significantly weaker transmission sources have to be employed in order to keep the detector dead time¹ within acceptable limits. The gain in sensitivity and scan time of a factor of four to five in three dimensions has to be given up almost completely when using transmission sources with only a fraction of the activity. One way to overcome this problem is to avoid using detectors close to the transmission sources, which would be most affected by high-dead-time loss, and to measure only the γ -rays that have passed through the patient and reached the opposite side of the gantry without coincidence detection. Since this approach eliminates the effect of “electronic collimation,” other means have to be used to determine the second coordinate of the “line of response.” This can be easily accomplished by using a radioactive point source rotating around the gantry aperture. In order to measure all lines of response necessary for three-dimensional attenuation correction, the point source also has to be stepped through all axial planes prescribed by the detector rings of the system. It has been shown that the transmission count rates that can be achieved with such a point source arrangement can be increased by more than a factor of 10 above conventional coincidence based transmission scans in three dimensions.

¹ The sensitivity loss at high count rates due to the event processing time in the γ -ray detector and the associated electronics.

6.4 Research Opportunities

To summarize, the following major areas of scientific or technological development appear to bear the greatest potential of enhancing the applications and clinical usefulness of PET:

- Cost-effective γ -ray detectors with high spatial, time, and energy resolution and the capability for measuring the angle and depth of interactions;
- Effective and fast means of acquiring low-noise transmission data;
- Mathematical techniques to take advantage of improved detector technologies, such as the possible inclusion of time-of-flight information into the reconstruction process;
- Fast three-dimensional reconstruction algorithms, in particular for fast dynamic or whole-body studies;
- Fast and quantitatively correct iterative reconstruction algorithms for two- and three-dimensional reconstructions;
- Inexpensive and powerful reconstruction processors to accommodate the above; and
- Fast real-time sorting electronics for data acquisition, and efficient data storage and handling capabilities for the vast amount of projection data in three-dimensional and whole-body studies.

6.5 Suggested Reading

1. Anger, H.O., Gamma-ray and positron scintillator camera, *Nucleonics* **21** (1963), 56.
2. Bailey, D.L., 3D acquisition and reconstruction in positron emission tomography, *Ann. Nucl. Med.* **6** (1992), 123–130.
3. Budinger, T.F., PET instrumentation, in *The Biomedical Engineering Handbook*, J.D. Bronzino, ed., CRC Press, Boca Raton, Fla., 1995, 1140–1150.
4. Casey, M.E., and Nutt, R., A multicrystal two-dimensional BGO detector system for positron emission tomography, *IEEE Trans. Nucl. Sci.* **33** (1986), 460–463.
5. Grootoink, S., Spinks, T.J., Kennedy, A.M., Bloomfield, P.M., Sashin, D., and Jones, T., The practical implementation and accuracy of dual window scatter correction in a neuro-PET scanner with the septa retracted, *IEEE Conf. Med. Imaging* **2** (1992), 942–944.
6. Kinahan, P.E., Rogers, J.G., Harrop, R., and Johnson, R.R., Three-dimensional image reconstruction in object space, *IEEE Trans. Nucl. Sci.* **35** (1988), 635–638.
7. Kübler, W., Ostertag, H., Hoverath, H., Doll, J., Ziegler, S., and Lorenz, W., Scatter suppression by using a rotating pin source in PET transmission measurements, *IEEE Trans. Nucl. Sci.* **35** (1988), 749–752.
8. Ollinger, J.M., and Johns, G.C., Model-based scatter correction in three dimensions, *IEEE Conf. Med. Imaging* **2** (1992), 1249–1252.
9. Ter-Pogossian, M.M., Phelps, M.E., Hoffman, E.J., and Mullani, N.A., A positron-emission transaxial tomograph for nuclear imaging (PETT), *Radiology* **114** (1975), 89–98.
10. Townsend, D.W., and Defrise, M., Image reconstruction methods in positron tomography, *Lectures Given in the Academic Training Program of CERN*, Report No. 93-02, CERN, Geneva, 1993.

Chapter 7

Ultrasonics

7.1 Introduction

Ultrasound, as currently practiced in medicine, is a real-time tomographic imaging modality. Not only does it produce real-time tomograms of scattering, but it can also be used to produce real-time images of tissue and blood motion, elasticity, and flow in the tissue (perfusion).

The most common form of imager is the B-scanner (the B merely stands for brightness). An ultrasonic pressure pulse about a microsecond long is launched into the tissue by a transducer consisting of an array of individually pulsed piezoelectric elements. This pulse is reflected from the various scatterers and reflectors within the tissue under investigation. The scattered pressure wave is detected by the transducer array and focused using electronic beam forming. The resulting signals are used to make an image that correlates to the scatterers and reflectors within the region from which the pressure pulse signal was reflected.

Because ultrasound is currently only a two-dimensional imaging modality, many investigators are researching multi-dimensional modalities such as three-dimensional imaging or are combining flow with scattering images in several dimensions. The combination of ultrasound imaging with therapeutics such as hyperthermia or drug injections, or with ultrasound ablation, is developing quickly. Intravascular and intracavitary imaging methods for both imaging and therapeutics are being investigated at a fast pace in both commercial development and basic research laboratories. Even ultrasonic microscopy, not covered here, is developing to surpass a resolution of 1 micron.

Technical advances in ultrasound supported by mathematics include computed tomography (inverse scattering), scatterer number density calculations (statistics), wave elastic tissue interaction (viscoelasticity), ferroelectric transducer development (ceramic physics), and wave equation modeling of ultrasound in viscoelastic materials such as tissue. The mathematical model behind ultrasound computed tomography is presented in section 14.1.3.

Mathematics and physics have greatly influenced the development of ultrasonic imaging, and many challenging problems from physics and the mathematical sciences remain to be solved. Beam forming in nonhomogeneous and usually nonisotropic materials such as biologic tissue is not at all well developed theoretically. Adaptive beam forming, which corrects for variations in refractive index within the imaged field, is a problem that is not solved. Inverse scattering is not a solved problem for geometries in which the sound is either traversing (forward scattering) or reflecting (backscattering) from the object. Acoustic models for the behavior of transducers with the complicated geometries of today's scanners are not well developed.

Following are some viewpoints on the various aspects of ultrasonic imaging.

7.2 Instrumentation

The instrumentation of ultrasonic imaging consists of the transducer, a beam former, and a signal analysis and display component.

7.2.1 Transducers

There are three important areas of development pertaining to transducers: (1) field distributions, (2) acoustics and vibration, and (3) electromechanical properties of piezoelectric and ferroelectric materials.

Field Distributions The seminal references in this area are, of course, related to the classic diffraction theory of Huygens and various analyses by Rayleigh, Kirchoff, and Sommerfeld. This theory is particularly applicable to monochromatic radiation and enables accurate calculations of the field distribution from ultrasound radiators under

certain conditions. Because it is limited to monochromatic ultrasonic energy, this approach is not well suited to wideband ultrasound imaging.

A method called the impulse response model was developed first by Oberhettinger in 1960 and subsequently refined by Stepanishen in the early 1970s into a linear systems model of beam propagation. This approach essentially transformed the problem from one of wave propagation to one of geometry. The great advantage of the impulse response approach was that analytic solutions existed for important classes of radiators such as spherical, rectangular, and annular apertures. It is now possible to make accurate calculations of wideband ultrasound field distributions for most relevant ultrasound transducer geometries.

Acoustics and Vibration The mathematical foundations were laid for acoustics and vibration analysis by Lord Rayleigh in his book *The Theory of Sound*, which was first published in 1896. From the transducer point of view, notable contributions were made by the Curie brothers, who discovered piezoelectricity in 1882, and by some of the great theoreticians such as A.E.H. Love, who made important contributions to the mathematical theory of elasticity. The theory of acoustics is well summarized in the book by Morse and Ingard listed in the suggested reading. The science of crystallography, exemplified in the work of Cady, had an important influence on the development of materials for ultrasound transducers. Similarly, the science of solid-state physics was critical in the development of an appreciation for the underlying physics of piezoelectricity and ferroelectricity. Recently, the classical theory of vibration has been extended by the development of finite-element models.

Electromechanical Properties of Ferroelectric Materials Two classes of models have been particularly useful in modeling electromechanical properties of ultrasound transducers. These are the Mason model, in which the transducer is modeled as a lumped element system, and the KLM (Krimholtz, Leedom, and Matthei) model, in which the transducer is modeled as a transmission line. In recent years, the KLM model has been particularly useful for transducer design and is currently used widely in this regard.

7.2.2 Ultrasonic Beam Forming

Phased array transducers contacting the surface of an object can be electronically scanned and focused to produce real-time images of internal structures. Unlike conventional radar phased arrays operating in the far field, electronic data processing of an ultrasound array emulates the diffraction integral of a lens focused to a particular point within the object. Early work in acoustical holography in the 1960s resulted in the first phased array ultrasonic systems in the early 1970s. Continuous progress on real-time processing has led to modern digital beam formers, which are summarized in the article by O'Donnell listed in the suggested reading. To date, all beam-forming systems have assumed a homogeneous distribution of the acoustical index of refraction (i.e., sound velocity). Imaging in highly inhomogeneous media, such as the human body, is severely degraded by this assumption.

7.2.3 Signal Processing

Ultrasound imaging has developed as an analog to radar imaging, but there are a number of areas where the two modes differ, and a more complete theory specific to the interaction of ultrasound and biological materials is required. The signal processing involved in forming images from A-mode data is fundamentally very simple and is based largely on linear (convolutional) models for sound transmission and reflection in tissue. Important early developments to form gray-scale, real-time images included the digital scan converter and the related software for preprocessing and postprocessing of images, along with signal processing schemes to achieve focusing and steering in ultrasonic arrays and to process the results. Currently a number of simple theories are used to model pieces of the imaging system: geometric optics for fixing transmit aperture (f-number), Fourier optics for apodization and control of side lobes, sampling theory for array design and system electronics, and statistical considerations for setting system dynamic range and color flow estimates. However, specific methods used by equipment manufacturers to form images appear to be fairly well guarded trade secrets.

Advances in real-time medical ultrasound imaging can only be achieved with a thorough understanding of the physics and physical mechanisms inherent to the imaging modality of interest, and the interactions of ultrasound with the body. The basis of any algorithm development must be an improved understanding of the underlying physics and physiological constraints of the problem. The literature is filled with examples of unsuccessful, ad hoc signal processing schemes tried on ultrasound images only because it was possible to do so. In years past, manufacturers have also periodically tried, and dropped, signal processing methods developed in an ad hoc manner.

Two fundamental effects that limit the quality of ultrasound imaging are speckle and phase aberration. Speckle is due to the coherent nature of the ultrasound signal, and phase aberration is caused by spatial variations of the propagation speed in tissue. The current state of affairs is to apply the multislice methods of x-ray computed tomography (CT) or magnetic resonance imaging (MRI) to form volumetric or surface rendered versions of tissue.

Ultrasound, however, is a real-time imaging method, and it is certainly going to be a challenge to make the multislice methods operate in real time. The development of novel data collection and processing methods for generating volumetric data sets should be given some attention. Methods that use a reduced amount of input data and rely on the brain's ability to sort out large amounts of data (perhaps combined with virtual reality display methods) would hold promise for developing practical three-dimensional real-time systems. Continuous wave techniques of computing motion (e.g., blood flow) are based on computing the Doppler shift in a reflected sinusoid that is performed by any of a number of well-known spectral estimation techniques now considered part of the signal processing literature.

A combined signal processor could have economic and performance advantages wherein it could take into account "speckles" in a two-dimensional space and track both blood movement and probe movement simultaneously. Tracking the probe movement provides a "compounding effect" to improve gray scale images and reduces noise in color flow imaging. Tracking the blood or tissue movement produces color Doppler and tissue motion imaging with direction, speed, and Doppler power information.

7.3 Scattering

Since the major sources of reflected and scattered waves are changes that occur at organ boundaries, even the crude early systems were able to produce images that correlated with anatomy and offered the first views of soft tissue structures that were invisible to simple transmission x-rays. Further refinements have used empirical techniques combined with medical knowledge to enhance the ability to image details, with image resolution as the main criterion. There is currently little or no ability to characterize tissue or tissue structures on any basis other than anatomical geometry or the flow patterns revealed by the Doppler effect. Since ultrasound instruments are widely used and distributed, there is considerable potential for improving diagnosis and treatment if better characterization can be done. The path to this goal is not clear. The benefits to be gained from a deeper understanding of the scattering process are difficult to predict, given the success of current methods and technology (see, for instance, the paper by Feleppa, Lizzi, and Coleman listed in the suggested reading).

All soft tissues except lung tissue have nearly the same composition, being mainly water with small amounts of protein. Therefore, the differences in bulk elastic properties are also small. This has led to the use of pulse-echo techniques for observing the weak scattered signals. The considerable energy losses in tissue also lead to frequency-dependent absorption. This absorption led to the early use of ultrasound as a source of therapeutic heat.

The imaging signal used by present-day clinical instruments is the product of the transmitted wave with scattering and absorption factors, and one factor cannot be measured without knowledge of the other. Laboratory measurement systems are needed to perform measurements free of this restriction. The frequency dependence of attenuation sets the resolution limit. Since the loss is inversely proportional to ultrasonic frequency, the size of the imaged region is proportional to wavelength. The size of the resolution cell is also proportional to wavelength. The images thus have a nearly constant number of resolution elements or pixels in depth and a constant relative resolution. Better resolution cannot be obtained at high frequencies without putting the transducers closer to the tissue to be imaged, and an understanding of absorption is quite important to our use of scattered waves.

Attenuation is thus a factor that influences the image when using scattered waves. Briefly, the peak pressure loss in tissue varies somewhere between the first and second power of frequency, depending on the tissue. The sound speed that is associated with this attenuation characteristic is almost constant with frequency, which makes simple pulse-echo imaging possible. There is evidence that attenuation is a property of the proteins, probably through vibration of the side groups of amino acids. It has not been possible to build up the attenuation characteristic of proteins from those of the constituent amino acids. This frequency dependence of attenuation at low frequencies violates causality, and the attenuation for unit wavelength must decrease at low and high frequencies. The only protein whose behavior has been studied to confirm this is hemoglobin, the attenuation of which must decrease below 50 kHz and above 300–500 MHz.

The propagation and scattering properties of liquids are easier to measure than are those of solids, and so most work has been done on blood. These results must apply to solid tissues to some degree, but more work is needed. The dispersion of sound speed in blood is very small and is in accord with predictions. The scattering is due to the formed elements, and their compressibility and density contributions have been separated. The density difference between red cells and plasma has been found to contribute to a "relative motion" absorption term at high frequencies.

These results for blood used the popular discrete model for scattering, together with the fluid properties of density and compressibility. Investigations correlated tissue properties with their water, lipid, and protein content. Further advances could result from applying the continuous model or the Lamé constants used in the theory of viscoelasticity. One problem in investigation is biological in nature: traditional models of molecular structure describe elements too small to affect scattering except in the large. At this level there are cells that are still much smaller than the wavelength, the current resolution-limiting factor. The structures that are of interest at this level (1000 to 100 microns) are not as well described in structure as are entire organs. There is considerable evidence that such

structures are important and contribute to the measured anisotropy of propagation and scattering. Future correlation of scattering with structure at these size levels could yield new insights.

The investigation of tissue properties at the levels suggested above could be done by further development and application of diffraction tomography using laboratory techniques. The future clinical application of these techniques could follow but would not be the initial reason for development. Because of the wavelength-based resolution limits of even experimental measuring systems, only the effective scattering cross section, or function averages over the resolution cell, can be measured. Complete computer-based algorithms for making such measurements corrected for system factors have been developed.

7.4 Ultrasonic Tomography

The fundamental groundwork for transmission tomography came from the mathematician Johann Radon in 1920. The inverse scattering problem in ultrasound, however, has not been solved, although accurate iteration methods have been developed. The problem can be broken into transmission tomography and reflection tomography, and this section deals only with the former. An approach using the Rytov approximation was developed by physicists Iwata and Nagata, and mathematicians have developed some methods using novel bases (see, for instance, the paper by Ball, Johnson, and Stenger in the suggested reading). Optical approaches have produced several ideas such as back propagation and diffraction tomography. Perturbation theory has been used as well (for instance, in the paper by Mueller, Kaveh, and Wade listed below).

Ultrasonic tomography has been applied to breast imaging, and a commercial product for that purpose is expected soon.

7.5 Research Opportunities

Many fruitful areas exist in ultrasonic imaging for further mathematical and physics research. The following is a list of areas of research that could produce a large impact on the field.

- Development of effective wideband models that include the physical acoustical properties of the tissues in which the beam is propagating, such as absorption and aberration-inducing effects. This is effectively the forward wave propagation problem.
- Further study of the dielectric and ferroelectric properties of materials, with particular emphasis on the development of new ferroelectrics with high dielectric constants, such as the relaxer ferroelectrics; of ferroelectric functions; and of the role of grain and domain boundaries in the piezo- and ferroelectric properties of materials.
- Development of four-dimensional finite-element models of single transducer elements and multielement arrays.
- Development of superbroadband transducers, array modeling, and effective modeling for multimode vibrations of transducer elements.
- Investigation of theories and development of faster beam-forming algorithms that will be flexible enough to allow adaptive corrections for phase and amplitude distortions attributable to both tissue and system effects. Development of algorithms that ensure that diagnostic information can be obtained from any one patient regardless of intervening tissue components.
- Development of statistical and/or adaptive methods to preserve image detail associated with high-level echoes (such as those that occur when surgical tools interact with tissue) and sharp boundaries around vessels and ducts while improving contrast detail associated with low-contrast or small objects. Development of contrast enhancement routines based on knowledge of underlying tissue types (both healthy and diseased) and their interaction with ultrasound. Investigation of genuine contrast improvement procedures through tissue characterization/parametric imaging (including texture analysis, tissue elasticity, and soon).
- Development of algorithms based on known differentiation between tissue types and healthy and diseased states.
- Improvement of methods for three-dimensional imaging capabilities. Development of better methods for displaying three-dimensional flow data in a format that is easily comprehensible. Development of image processing routines that overcome the specific problems of three-dimensional ultrasound (e.g., acquisition issues, coherent image formation (speckle), inhomogeneous media, non-transmission modes) and thus allow optimal utilization of existing three-dimensional reconstruction techniques. Development of methods for region segmentation as a preprocessing step to measurements of both area and volume, and application of three-dimensional image data display methods and automatic methods for measuring areas and volumes.

- Development of future imaging systems that use scattered waves more effectively. The present pulse-echo systems use a fractional bandwidth greater than 50%. Because the waves are basically oscillatory and coherent, the images show “speckle,” the multiplicative random noise noted in laser images. If a wider bandwidth could be obtained by extending the response to low frequencies, a speckle-free system could result.
- Development of physics for the inverse problem to include the complete viscoelastic wave equation and its solutions for both transmission and reflection tomography.

7.6 Suggested Reading

1. Ball, J., Johnson, S.A., and Stenger, F., Explicit inversion of the Helmholtz equation for ultra-sound insonification and spherical detection, in *Visualization and Characterization*, K.Y. Wang, ed., Plenum Press, New York, 1980, 451–461 (volume 9 in Acoustical Imaging series).
2. Borup, D.T., Johnson, S.A., Kim, W.W., and Berggren, M.J., Nonperturbative diffraction tomography via Gauss-Newton iteration applied to the scattering integral equation, *Ultrasonic Imaging* **14** (1992), 69–85.
3. Cady, W.G., *Piezoelectricity: An Introduction to the Theory and Applications of Electromechanical Phenomena in Crystals*, Dover Publications, New York, 1964.
4. Devaney, A.J., A filtered backpropagation algorithm for diffraction tomography, *Ultrasonic Imaging* **4** (1982), 336–350.
5. Feleppa, E.J., Lizzi, F.L., and Coleman, D.J., Ultrasonic analysis for ocular tumor characterization and therapy assessment, *News Physiol. Sci.* **3** (1988), 193–197.
6. Fredrich, W., Kaarmann, H., and Lerch, R., Finite element modelling of acoustic radiation from piezoelectric phased array antennas, *Proc. Ultrasonic Symp.* **1** (1990), 763.
7. Iwata, K., and Nagata, R., Calculation of refractive index distribution from interferograms using Born and Rytov’s approximation, *Jpn. J. Appl. Physiol.* **14** (Suppl.) (1975), 379–383.
8. Kirchoff, G., Zur theorie der lichstrahlen, *Wiedemann Ann.* **18** (1883), 663.
9. Krimholtz, R., Leedom, D.A., and Matthei, G.L., New equivalent circuits for elementary piezoelectric transducers, *Elect. Lett.* **6** (1970), 398–399.
10. Leedom, D.A., Krimholtz, R., and Matthei, G.L., Equivalent circuits for transducers having arbitrary even or odd symmetry piezoelectric excitation, *IEEE Trans. Sonics Ultrason.* **SU-18** (1971), 128–141.
11. Love, A.E.H., *Mathematical Theory of Elasticity*, Dover, New York, 1944.
12. Mason, W.P., *Piezoelectric Crystals and Their Application to Ultrasonics*, H.M. Stationery Office, London, 1950.
13. Morse, P.M., and Ingard, K.U., *Theoretical Acoustics*, reprinted by Princeton University Press, Princeton, N.J., 1986.
14. Mueller, R.K., Kaveh, M., and Wade, G., Reconstructive tomography and applications to ultrasonics, *Proc. IEEE* **67** (1979), 567–587.
15. Oberhettinger, F., On transient solutions of the “baffled piston” problem, *J. Res. Natl. Bur. Stand., Sect. B* **65B** (1961), 106.
16. O’Donnell, M., Applications of VLSI circuits to medical imaging, *Proc. IEEE* **76** (1988), 1106–1114.
17. Rayleigh, J.W.S., *The Theory of Sound*, MacMillan, New York, 1896.
18. Sommerfeld, A., Mathematische theorie der diffraction, *Math. Ann.* **47** (1896), 317.
19. Stepanishen, P.R., Transient radiation from pistons in an infinite planar baffle, *J. Acoust. Soc. Am.* **49** (1971), 1629–1638.

Chapter 8

Electrical Source Imaging

8.1 Introduction

Electrical source imaging (ESI) is an emerging technique for reconstructing brain or cardiac electrical activity from electrical potentials measured away from the brain or heart. The concept of ESI is to improve on electroencephalography (EEG) or electrocardiography (ECG) by determining the locations of sources of current in the body from measurements of voltages. ESI could improve diagnoses and guide therapy related to epilepsy and heart conduction abnormalities through its capability for monitoring the effects of drugs or for locating an electrical abnormality that is to be removed.

Differences in potential within the brain, heart, and other tissues reflect the segregation of electrical charges at certain locations within these three-dimensional conductors as nerves are excited, causing cell membrane potentials to change. While the potential measured at some distance from an electrical charge generally decreases with increasing distance, the situation is more complex within the body; generators of the EEG, for example, are not simple point-like charge accumulations but rather are dipole-like layers. Moreover, these layers are convoluted and enmeshed in a volume conductor with spatially heterogeneous conductivity. The particular geometry and orientation of these layers determines the potential distribution within or at the surface of the three-dimensional body.

The classical approach to studying brain electrical activity involves recognizing patterns in a set of waveforms showing voltage as a function of time, acquired from about 20 electrodes placed on the scalp. While frequency analysis methods can indicate probable Alzheimer's disease by the abnormal distribution of spatial frequency bands (see Fig. 1.1), the true distribution of neuronal activity, knowledge of which could lead to more refined diagnoses, is masked or blurred by the conducting tissue layers between the central cortex and the electrodes.

Cardiac electrical activity is likewise spatially complex, and involves the propagation of excitation wave fronts in the heart. Standard electrocardiographic techniques such as electrocardiography (ECG) and vectorcardiography (VCG) are very limited in their ability to provide information on regional electrical activity and to localize bioelectrical events in the heart. In fact, VCG lumps all cardiac wave fronts into a single dipole located at the "center" of the heart and known as the "heart vector." Traditional ECG and VCG employ a small number of electrodes (6 or 12) to measure potentials from the body surface, and the patterns of electrical activity cannot give the information required for characterizing the electrical activity of the heart. Non-invasive electrocardiography requires simultaneous recordings of electrical potential from 100 to 250 torso sites in order to map the body surface potential. These body surface potential maps (BSPMs) reflect the regional time course of electrical activity of the heart, information that is important for clinical treatment; an example is localization of the accessory atrioventricular pathway in Wolff-Parkinson-White syndrome prior to surgical dissection or catheter ablation. Figure 8.1 shows a 180-electrode BSPM vest. The computer-controlled data acquisition system permits simultaneous recording from all 180 sites every millisecond throughout the cardiac cycle.

Body surface potential distribution is a very low resolution projection of cardiac electrical activity, and details of regional electrical activity in the heart cannot be determined merely from visual inspection of the BSPMs. A mathematical method of reconstructing endocardial potentials is greatly needed. The endocardial surface of the heart is accessible through catheterization procedures. Recently, multielectrode probes have been developed with 65–100 evenly distributed electrodes on their surface that do not occlude the ventricular cavity and thus can be used in the catheterization laboratory for clinical and basic medical science studies.

Figure here Figure 8.1. Left: The multielectrode (180 or 250) mapping vest for obtaining body surface potential maps. Right: The geometry used in the definition of the Cauchy problem in electrocardiography.

Since the probe electrodes are not in direct contact with the endocardium, the measured potentials can provide only a low-resolution reflection of the endocardial potentials, smoothed out by the high conductivity of the intracavitary blood. A mathematical procedure for computing the distribution of endocardial potential from the measured potentials over the catheter surface would, therefore, enhance resolution and permit the localization of electrical cardiac events during electrophysiology studies. With this procedure, detailed endocardial maps could be obtained in the catheterization laboratory on a beat-by-beat basis. Such a procedure would permit studies of arrhythmogenesis in intact animals and patients, including studies of mechanisms, dynamic changes (initiation, termination), and the effects of drugs. It would also permit an accurate determination, in the catheterization laboratory, of the site of origin of an arrhythmia prior to surgical or catheter ablation. This information could be obtained in a few minutes, and the recording time could be limited to one cardiac cycle (with no need for sequential recordings). As catheter ablation techniques are developed, an accurate determination of the arrhythmogenic site in the catheterization laboratory is the limiting factor to the successful implementation of the ablation procedure.

Both cerebral cortex and epicardial reconstructions require multielectrode mapping of electric potential on a closed surface (the scalp and torso, respectively).

8.2 Outline of ESI Reconstruction Methods

The reconstruction procedures described above are characterized mathematically as solutions to the inverse problem of electroencephalography and electrocardiography. The mathematical approach to both reconstruction procedures is the same and is outlined below with a focus on the torso-epicardium inverse problem. The same principles apply to the scalp-cortex problem of electroencephalography.

The inverse problem in electrocardiography, formulated in terms of epicardial potentials, constitutes the following Cauchy problem (see also section 14.1.10):

$$\begin{aligned} \nabla^2 \phi(x) &= 0 \quad \text{in } \Omega \\ \partial \phi(x) / \partial n &= \Gamma_1 \quad \phi(x) = Z(x) \quad \text{on } \Sigma \subseteq \Gamma_1 \\ \text{Find} \quad \phi(x) &\quad \text{on } \Gamma_\Omega. \end{aligned}$$

The geometry is defined in Figure 8.1 (right), $\phi(x)$ is the potential in Ω , $Z(x)$ is the (known) potential distribution on the torso, and $\partial \phi(x) / \partial n$ is the normal derivative of $\phi(x)$ on Γ_1 , which has to be zero since the normal component of current at the torso surface is zero. (The torso is in air, an insulating medium that does not support current flow.)

While a unique solution to this problem exists, the problem is ill posed in the sense that small perturbations in the data (noise) result in unbounded errors in the solution. This property implies that a straightforward solution is not possible because the data are never available without a certain degree of noise. Large deviations from the actual solution may result from inaccuracy of the data measurement, incomplete knowledge of the potential over the entire body surface, or an inaccurate description of the inhomogeneous torso.

The numerical solution for the inverse problem requires two steps. The first step is the construction of a mathematical expression describing the potential field throughout the torso volume conductor, particularly on the body surface, as a function of epicardial potential data (the forward problem). This has generally been done with either the finite-element method (FEM) or the boundary-element method (BEM). The second step is the inversion of that expression to enable the reconstruction of epicardial potentials from the measured torso potential data (the inverse problem). The following describes the methodology for each step.

8.2.1 Forward Problem

The BEM is a subset of the FEM, where the surface bounding the volume is divided into two-dimensional elements; each surface element has a basis function describing the distribution of electrical potential. While both the conductivity and the geometry of the volume are taken into account in the solution of the problem, only the relationships among the various regions of the surfaces appear in the final formulation. A major advantage of the BEM as compared to the FEM is that it involves fewer nodes (only the surface) and hence less overhead in construction and in modifying the mesh. Its major disadvantage is that its matrix is full and therefore requires more computer memory. With the capacity of the modern computer, however, this characteristic is not of great concern.

The BEM is derived from Green's second theorem:

$$\int_V \left[\phi \nabla^2 \left(\frac{1}{r} \right) - \left(\frac{1}{r} \right) \nabla^2 \phi \right] dV = \int_S \left[\phi \frac{\partial(1/r)}{\partial n} - \frac{1}{r} \frac{\partial \phi}{\partial n} \right] dS.$$

In this expression the surface S includes the epicardial surface S_E and the torso surface S_T , and V is the homogeneous volume enclosed between the surfaces S_E and S_T . The scalar potential ϕ satisfies Laplace's equation $\nabla^2 \phi = 0$ within

Figure here Figure 8.2. Example of a completely non-invasive computation. Early QRS (110 ms into the cardiac cycle, right ventricular breakthrough). (A) Measured epicardial potentials. (B) Inverse-recovered epicardial potentials obtained by Tikhonov zero-order regularization. (C) Computation of the epicardial distribution with temporal regularization using inverse-computed (Tikhonov zero-order) epicardial potential distribution from the next time frame (112 ms) for the estimate vector \mathbf{P} . Anterior potentials are on the left; posterior potentials are on the right. Maxima (+) and minima (−) values, contour lines, and intervals are in microvolts. Solid contour lines indicate positive potentials; broken contour lines indicate negative potentials. RE: relative error. Arrows in the maps identify the anterior minimum that reflects right ventricular epicardial breakthrough. (Modified from Oster, H.S., and Rudy, Y., The use of temporal information in the regularization of the inverse problem of electrocardiography, *IEEE Trans. Biomed. Eng.* **BME-39** (1992), 65–75.)

the volume; hence, the second term on the left-hand side is zero. The variable r is the distance between a field point p and a source point q within that volume; $1/r$ is the Green's function or fundamental solution satisfying Poisson's equation at a point. $\partial/\partial n$ is the normal derivative.

8.2.2 Inverse Problem

The inverse problem is formulated by dividing the brain cerebral cortex or epicardial surface S_E into M_E triangular elements and the torso surface S_T into M_T triangular elements. After manipulation and regrouping of terms, the following matrix equation is obtained:

$$V_T = AV_E,$$

where V_E is the vector of cerebral cortex or epicardial potentials, V_T is the vector of scalp or torso potentials, and A is the transfer matrix between heart and torso that depends on anatomic and electrical properties of heart-torso geometry. The inverse problem requires an inversion of A to give V_E in terms of V_T .

Inversion is performed with a least-squares approach, minimizing the residual vector with respect to V_E , leading to the solution

$$V_E = (A^*A)^{-1}A^*V_T,$$

where * refers to transpose.

Because eigenvalues of the matrix A are very small, $(A^*A)^{-1}$ greatly magnifies whatever it multiplies. If V_T contained the exact values of scalp or torso electric potential, this last equation would give an excellent estimate of the solution V_E . However, any inevitable error term becomes greatly magnified, giving an erroneous (and sometimes unbounded) "solution." Techniques of regularization must therefore be used to minimize the effects of such error by imposing constraints on the solution. Many of the regularization methods produce similar results and even reduce to similar forms for certain parameter values. The Tikhonov regularization method (see also section 14.4.1) has been found to give good results in this electrical mapping problem. V_E is obtained by minimizing the objective function

$$\min[\|AV_E - V_T\|^2 + t\|RV_E\|^2],$$

where t is a regularization parameter and R is either the unit matrix I (Tikhonov method of order zero), the surface gradient operator G (Tikhonov method of first order), or the surface Laplacian operator L (Tikhonov method of second order). The solution now becomes

$$V_E = (A^*A + tR^*R)^{-1}A^*V_T.$$

8.2.3 Temporal Regularization

The electrical mapping problem is also characterized by the fact that brain and cardiac activation is a continuous process in time. Incorporation of the continuous time progression into the regularization procedure has been done using Twomey regularization. Because of the continuity of the activation sequence in time, it is reasonable to expect that an estimate of the solution at one time can be made from the information at other time frames, and that this will improve the estimate and hence the quality and accuracy of the inverse-reconstructed heart potential map (e.g., Fig. 8.2).

Instead of imposing constraints on the magnitude of the solution or on its derivatives, the Twomey method minimizes the difference between the solution V_E and some estimate of the solution, P . The objective function becomes

$$\min_{V_E}[\|AV_E - V_T\|^2 + t\|V_E - P\|^2].$$

The solution to this new objective function is

$$V_E = (A^*A + tI)^{-1}(A^*V_T + tP).$$

Of course, the objective function reduces to the Tikhonov zero-order objective function when $P = 0$.

The estimate P of epicardial potentials at a given time frame can be constructed using information from other time frames, and the optimal method of generating this estimate from historical and future time frames is yet to be determined.

8.3 Research Problems and Opportunities

ESI is a promising modality, but the number and quantitative precision of the brain or torso measurements and the degree of understanding of the conductive paths may be insufficient to allow an accurate solution of the Cauchy problem. The following research areas need further development before the potential of ESI can be realistically gauged.

- The inverse computation of cerebral cortex and epicardial potentials requires knowledge of the geometries and conductivities of intervening tissues. It may not be practical, in an experimental or clinical setting, to even measure the position of each body surface electrode. Some convenient strategy of individual-specific anatomic measurement (e.g., by MRI) needs to be developed so that the true geometry and inhomogeneous conductivity can be incorporated into the reconstructions of potentials.
- Resolution needs to be improved to the point that details of conductive disturbances (e.g., areas of slow conduction in the heart) can be detected and localized. This could be achieved by modifying the mathematical formulation to incorporate higher-resolution temporal and spatial information into the regularization procedure. This would include development of fast algorithms for computing and displaying reconstructed information at close to real time.
- The accuracy of the inverse procedure is dependent on the number and position of the surface electrodes. Methods are needed to characterize the optimal number and positions for expected noise conditions.
- Although potentials can be reconstructed with good accuracy using Tikhonov inversion techniques, a significant improvement is achieved when the time progression is incorporated into the regularization procedure for the cardiac inverse problem. These techniques could be extended to allow similar improvements in studies of the brain and of heart activation.

8.4 Suggested Reading

1. Balish, M., and Muratore, R., The inverse problem in electroencephalography and magnetoencephalography, in *Magnetoencephalography*, Susumu Sato, ed., Raven Press, New York, 1990.
2. Gisser, D.G., Isaacson, D., and Newell, J.C., Electric current computed tomography and eigenvalues, *SIAM J. Appl. Math.* **50** (1990), 1623–1634.
3. Gloor, P., Neuronal generators and the problem of localization in electroencephalography: Application of volume conductor theory to electroencephalography, *J. Clin. Neurophysiol.* **2** (1985), 327–354.
4. Law, S.K., Nunez, P.L., and Wijesinghe, R.S., High-resolution EEG using spline generated surface laplacians on spherical and ellipsoidal surfaces, *IEEE Trans. Biomed. Eng.* **40** (1993).
5. Messinger-Rapport, B.J., and Rudy, Y., Regularization of the inverse problem in electrocardiography: A model study, *Math. Biosci.* **89** (1988), 79–118.
6. Messinger-Rapport, B.J., and Rudy, Y., Non-invasive recovery of epicardial potentials in a realistic heart-torso geometry: Normal sinus rhythm, *Circ. Res.* **66** (1990), 1023–1039.
7. Oster, H.S., and Rudy, Y., The use of temporal information in the regularization of the inverse problem of electrocardiography, *IEEE Trans. Biomed. Eng.* **BME-39** (1992), 65–75.
8. Pilkington, T.C., and Plonsey, R., eds., *Engineering Contributions to Biophysical Electrocardiography*, IEEE, New York, 1982.

9. Rudy, Y., and Oster, H.S., The electrocardiographic inverse problem, in *High Performance Computing in Biomedical Research*, T.C. Pilkington, B. Loftis, J.F. Thompson, S.L.-Y. Woo, T.C. Palmer, and T.F. Budinger, eds., CRC Press, Boca Raton, Fla., 1993, 135–155.
10. Webster, J.G., ed., *Electrical Impedance Tomography*, Adam Hilger, New York, 1990.

Chapter 9

Electrical Impedance Tomography

9.1 Introduction

Electrical impedance tomography (EIT) uses low-frequency electrical current to probe a body; the method is sensitive to changes in electrical conductivity. By injecting known amounts of current and measuring the resulting electrical potential field at points on the boundary of the body, it is possible to “invert” such data to determine the conductivity or resistivity of the region of the body probed by the currents. This method can also be used in principle to image changes in dielectric constant at higher frequencies, which is why the method is often called “impedance” tomography rather than “conductivity” or “resistivity” tomography. However, the aspect of the method that is most fully developed to date is the imaging of conductivity/resistivity. While EIT methods have not yet gained a significant foothold in the medical imaging community, they have been shown to work well in both geophysical and industrial settings and, therefore, it is possible that future medical imaging applications may follow rather rapidly from the advances made for other applications.

9.2 Comparison to Other Modalities

There is a formal mathematical analogy between EIT and x-ray computed tomography (CT), since in either case data must be processed to produce the desired image of interior structure and, furthermore, the imaging is often performed on two-dimensional slices through the body. EIT uses diffusion of current to deduce conductivity distribution, unlike methods such as magnetic resonance imaging (MRI), CT, positron emission tomography (PET), and single photon emission computed tomography (SPECT).

EIT is expected to have relatively poor resolution compared to MRI, CT, PET, and SPECT. Resolution is largely controlled by the number of electrodes that can be reasonably attached simultaneously to a patient. Schemes used to date have normally used electroencephalogram-style electrodes, relatively few in number and large in size. It is not yet known whether special belts or headbands having large numbers of small electrodes designed especially for EIT applications might considerably improve this situation. But, regardless of such possible advances in technology, it is not anticipated that EIT will ever “outresolve” methods like x-ray CT.

At the present time, EIT is the only method known that images electrical conductivity, although MRI and electromagnetic methods also have some potential to measure conductivity. So, for applications requiring knowledge of the distribution of this parameter through a body, EIT will continue to be an important method to consider for medical imaging, regardless of its resolving power.

On the other hand, EIT has some very attractive features. The technology for doing electrical impedance imaging is safe and inexpensive, and therefore could be made available at multiple locations (for example, at bedside) in hospitals. At the low current levels needed for this imaging technique, the method is not known to cause any long-term harm to the patient, and therefore could be used to do continuous (or frequent, but intermittent) monitoring of bedridden patients. Technology for acquiring data, and algorithms for inverting that data to produce images of conductivity/resistivity, have been developed to the point that real-time imaging could become routine today using a graphics workstation as the computing platform.

9.3 Present Status of EIT and Limitations

The impedance imaging problem is nonlinear and extremely ill posed, which means that large changes in interior properties can result in only small changes in the measurements. This implies that making a high-resolution image

would require extremely accurate measurements, which in turn makes the problem of designing appropriate electronics very challenging.

At present there are two main approaches to the problem. The first, called the applied potential tomography (APT) system, was developed by Barber and Brown in Sheffield, England. Surrounding the body by a ring of 16 electrodes, they used a single current generator to pass current between a pair of adjacent electrodes and measure the voltage on all the other electrodes; they then connected the current generator to the next pair of adjacent electrodes, and so on, until all adjacent pairs had served as “driver pairs.” This approach has the advantage of simplicity of design, but its resolution is intrinsically limited; it is also sensitive to small errors in electrode placement. Because of this sensitivity, the APT system cannot make images of the electrical parameters themselves, but can only display changes. This system has been used in studies of various physiological processes, including gastric emptying, blood flow in the thorax, head, and arm, and pulmonary ventilation.

The second approach to impedance imaging is the adaptive current tomograph (ACT), developed by a team at Rensselaer Polytechnic Institute. These systems use 32 or 64 electrodes, each with its own programmable current generator. Although this design is more complicated, it overcomes many of the limitations of the APT design: ACT systems are optimal in the sense that they extract the maximum possible information, and they are less sensitive to electrode placement errors. ACT systems have produced images of the electrical conductivity and permittivity in the human thorax, and have been used in breast studies.

The amount of research in EIT is increasing rapidly. Most of this research is taking place in Europe, where researchers have worked on topics ranging from system design and reconstruction algorithms to clinical applications. Most of the U.S. research, other than that of the Rensselaer group, has been concerned with mathematical aspects of the reconstruction problem; see section 14.1.5.

9.4 Research Opportunities

- Development of a realistic model that describes the multiple path problem, and development of mathematical methods of tomography that respect the physics of multipath incoherent diffusion.
- Clinical research performed to determine the role of impedance imaging for applications such as determination of cardiac output, monitoring for pulmonary edema, and screening for breast cancer.

9.5 Suggested Reading

1. Barber, D.C., Electrical impedance tomography, in *The Biomedical Engineering Handbook*, J.D. Bronzino, ed., CRC Press, Boca Raton, Fla., 1995, 1151–1164.
2. Barber, D.C., Brown, B.H., and Seaper, A.D., Applied potential tomography: Possible clinical applications, *Clin. Phys. Physiol. Meas.* **6** (1985), 109–121.
3. Berryman, J.G., Convexity properties of inverse problems with variational constraints, *J. Franklin Inst.* **328** (1991), 1–13.
4. Isaacson, D., Distinguishability of conductivities by electric current computed tomography, *IEEE Trans. Med. Imaging* **MI-51** (1986), 91–95.
5. Isakov, V., On uniqueness of recovery of a discontinuous conductivity coefficient, *Commun. Pure Appl. Math.* **41** (1988), 865–877.
6. Webster, J.G., *Electrical Impedance Tomography*, Adam Hilger, Bristol, 1990.
7. Yorkey, T.J., Webster, J.G., and Tompkins, W.J., Comparing reconstruction algorithms for electrical impedance tomography, *IEEE Trans. Biomed. Eng.* **BME-32** (1987), 843–852.

Chapter 10

Magnetic Source Imaging

10.1 Introduction

Ion currents arising in the neurons of the heart and the brain produce magnetic fields outside the body that can be measured by arrays of SQUID (superconducting quantum interference device) detectors placed near the chest or head; the recording of these magnetic fields is known as magnetocardiography (MCG) or magnetoencephalography (MEG). Magnetic source imaging (MSI) is the reconstruction of the current sources in the heart or brain from these recorded magnetic fields. These fields result from the synchronous activity of tens or hundreds of thousands of neurons.

Both magnetic source imaging and electrical source imaging (see Chapter 8) seek to determine the location, orientation, and magnitude of current sources within the body. The magnetic field at the surface is most strongly determined by current sources directed parallel to the surface, but the electrical potentials are determined by current sources directed perpendicular to the surface. Other than the signal distortion from the heterogeneity of tissue conductivity, there is no clear physical reason that the clinical information produced by biomagnetic measurements could not as well be obtained from electrical potential mapping. An advantage of MSI over ESI is that all body tissues are magnetically transparent and the magnetic fields propagate to the surface without distortion. The electrical potentials at the surface, on the other hand, are distorted by variations in conductivity within the body; this is especially true in the head, where the low conductivity of the skull both distorts and hides the electrical activity of the brain. A disadvantage of MSI is that the need for cryogenic cooling and a magnetically shielded room makes the procedure cumbersome with the present technology.

Biomagnetism offers a tool to study processes where electrical function is important. Promising results have been obtained in the fields of cardiology and epilepsy. Sites of origin for heart arrhythmias (e.g., the accessory path for Wolff-Parkinson-White syndrome) can be identified by electrical activity in abnormal anatomical locations. Non-invasive localization allows treatment by guiding an ablation catheter directly to the correct site. MSI can also be used in the surgical treatment of intractable epilepsy to locate the epileptogenic focus and functional areas of the brain that must be conserved during surgery. A potential use in neuroscience is the spatial and temporal study of functional processing areas in the brain in response to auditory, visual, and somatosensory stimuli. Biomagnetism also offers a research tool for studying schizophrenia as well as Parkinson's and Huntington's syndromes. The function of peripheral nerves can be studied, and prenatal magnetocardiography can be carried out.

A major strength of MSI is that it can resolve events separated by milliseconds, whereas other methods such as functional magnetic resonance imaging (fMRI), magnetic resonance spectroscopy (MRS), positron emission tomography (PET), and single photon emission computed tomography (SPECT) have time resolutions of seconds to many minutes, depending on the information sought. A weakness of MSI is that any magnetic field distribution on the surface of the head can be explained by an infinite number of current distributions inside the head. Thus, a successful source analysis is dependent on the availability of additional information suitable for constraining the inverse problem to be solved.

Because of their extreme weakness, on the order of femto-teslas, biomagnetic fields require sensitive sensors. Up to now only superconducting quantum interference devices (SQUIDS) have been able to achieve the required sensitivity below $10 \text{ fT}/\sqrt{\text{Hz}}$. In order to separate the biomagnetic fields from much larger external interferences, screening chambers made from soft magnetic and conductive material are necessary. The two technical areas most in need of development are simplification of the detector array and creation of methods to remove interference without the use of an expensive (e.g., \$500,000) enclosure. An active screen applying a closed control loop with a sensor for the interfering field and compensating current loops could be an attractive alternative to the passive screen.

SQUIDS based on high-temperature superconductor technology could operate with inexpensive liquid nitrogen.

Current technology has already led to devices with a sensitivity of about $100 \text{ fT}/\sqrt{\text{Hz}}$, which is still an order of magnitude above that achieved with “classic” superconductivity. It is likely, however, that further improvements are possible.

10.2 Mathematical Considerations

The inverse problem of calculating the current density $\mathbf{J}(\mathbf{p})$ within the body as a function of position \mathbf{p} from the magnetic field $\mathbf{B}(\mathbf{r})$ is defined by the Biot-Savart law,

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{\mathbf{R}^3} \frac{\mathbf{J}(\mathbf{p}) \times (\mathbf{r} - \mathbf{p})}{\|\mathbf{r} - \mathbf{p}\|^3} d^3\mathbf{p}.$$

See also section 14.1.9.

The current distribution within the body consists of two components: a non-ohmic primary (or impressed) current that is directly related to the physiological processes, and a volume current induced by the primary current sources in the conductive body tissue. Since the body is moderately conductive, the current sources being imaged generate “volume currents” within the body that depend on the body geometry and the distribution of tissue conductivity. These volume currents generate their own magnetic fields and thereby modify the fields generated by the primary current sources themselves. Accounting for the effects of these volume currents improves the accuracy of the reconstruction. In some simple geometries, it can be shown that the volume currents have no effect on the magnetic field outside the body and thus can be ignored; this is the usual assumption for contemporary implementations of MSI in commercial and research instruments. But for realistic body geometries, the effect of the volume currents must be included for an accurate reconstruction. General models allowing anisotropic conductivity have been studied using finite-element or finite-difference methods, but at present these techniques require much computation time and are not yet practical. Given a primary current distribution $\mathbf{J}^p(\mathbf{p})$, the electrical potential ϕ can be computed by solving Poisson’s equation

$$\nabla \cdot (\sigma \nabla \phi) = \nabla \cdot \mathbf{J}^p$$

with the boundary condition that $\mathbf{n} \cdot \nabla \phi = 0$ on the body surface. The conductivity σ must be known as a function of position and may be determined by a prior magnetic resonance scan. Then the total current \mathbf{J} is given by Ohm’s law

$$\mathbf{J} = \mathbf{J}^p - \sigma \nabla \phi.$$

These equations can be solved by various finite-element and boundary-element methods.

Given the magnetic field as a function of position, the magnetic flux ψ_m threading a coil of area A_m is simply the surface integral

$$\psi_m = \int_{A_m} \mathbf{B} \cdot d\mathbf{A}.$$

The fact that reasonable localization results for dipolar sources can often be obtained with extremely simplified models (e.g., a homogeneous sphere) suggests that relatively simple volume conductor models (e.g., a piecewise homogeneous volume conductor) might suffice to provide localization accuracies of the order of millimeters. However, volume conductor models with more complicated properties (e.g., anisotropic conductivities) represent no major problems.

10.3 Source Models

Because contemporary SQUID detector arrays do not provide sufficient information to estimate the underlying source configuration in a unique way, any interpretation of magnetic field data requires certain assumptions. In most studies the assumption is made that only a single source is active at a given time point and that the spatial extent of the source is small compared to the distances between source and measurement locations. The resulting model is the equivalent current dipole. Even though the assumption of a single dipole is generally an oversimplification, a source analysis based on this model can provide at least an estimate of the center of the source and of the mean activity in the source. A straightforward extension of the single dipole model is the multidipole model, with two or more spatially separated equivalent current dipoles. Model simulations have shown that two simultaneously active dipoles can be separated only if the spacing of the dipoles is comparable to the distances between the dipoles and measurement locations. Higher spatial resolution can be expected if time intervals rather than single time points are considered.

The single-source model assumes that the observed fields are generated by a few point sources (typically current dipoles) and seeks to determine the position, orientation, and magnitude of these dipoles from the observed fields.

That is, the unknown current distribution is written in the form

$$\mathbf{J}^p(\mathbf{r}) = \sum_{n=1}^N a_n \mathbf{J}_n^p(\mathbf{r}; \mathbf{p}_n, \mathbf{q}_n),$$

where the amplitude a_n , position \mathbf{p}_n , and orientation \mathbf{q}_n of each source are unknowns to be determined in the reconstruction. Since the fields depend nonlinearly on the dipole position and orientation, iterative optimization algorithms are required; such iterative algorithms are expensive in computer time and often find local rather than global minima. These methods work reasonably well for source models with one or perhaps two dipoles but rapidly become more expensive to compute and less reliable as the number of dipoles increases.

An alternative to single and multiple dipoles is a diffuse current distribution, which can be discretized in terms of a large number of homogeneously distributed current dipoles. Models of this type have the advantage that the source structure need not be defined in advance. However, the number of model parameters is generally much larger than the number of measurement points, and so a solution has to be selected among an infinite number of possible solutions. Choosing the current density with the smallest root mean square (the “minimum norm solution”) gives a three-dimensional image that depends on the choice of regularization technique or source model.

This approach assumes that the observed fields are generated by a distributed current source and then approximates that source as a weighted sum of elementary sources that are fixed in position; then it is necessary only to find the weights (or source amplitudes) that best approximate the observations. That is, the unknown current distribution is written in the form

$$\mathbf{J}^p(\mathbf{r}) = \sum_{n=1}^N a_n \mathbf{J}_n^p(\mathbf{r}),$$

where the position \mathbf{p}_n and orientation \mathbf{q}_n are fixed in advance, leaving only the amplitudes a_n to be determined in the reconstruction. The elementary sources are often dipoles, but other distributed sources may also be used. Fixing the source positions reduces the problem to one in which there is a unique global minimum. The observed magnetic flux measurements are linear in the unknown amplitudes, or

$$\psi = \mathbf{F}\mathbf{a},$$

where ψ is a vector of magnetic flux measurements, \mathbf{a} is a vector of the unknown source amplitudes, and \mathbf{F} is the forward transfer matrix connecting them. Iterative solution is not required, and the solution can be rapidly computed by numerical linear algebra software.

In sum, three paths can be pursued: (1) a single source can be reconstructed with rather high spatial resolution; (2) distributed current densities can be reconstructed with poor spatial resolution; or (3) several independent single sources can be identified from the time course of the magnetic field and reconstructed similar to a single source. All three methods require a model of the conductivity distribution of the head or torso under investigation. If this is taken from individual three-dimensional medical images, significant computational efforts are required.

10.4 Resolution

The limited number of detectors and the low signal-to-noise ratio means that MSI has poorer spatial resolution (1–2 cm) than do many other medical imaging modalities; furthermore, depending on the source strength, sources in the brain must be within several centimeters of the surface to be reliably detected; sources in the heart are stronger and deeper sources can be detected there. The temporal resolution of MSI, however, is milliseconds or tens of milliseconds, which makes it attractive for neuroscience studies.

A current line of research is to improve the apparent spatial resolution of MSI by using prior knowledge of the anatomy obtained from magnetic resonance imaging (MRI) or computed tomography (CT) to constrain the magnetic source reconstruction. One approach is simply to superimpose reconstructed dipole solutions on a magnetic resonance image registered to the same coordinate system; dipole locations not consistent with the anatomy may then be rejected as artifacts. This has been done with dipole point sources and is useful in interpreting the reconstruction.

A more sophisticated approach is to constrain the reconstruction by permitting sources only where they are consistent with the anatomy. For example, sources in the skull and in the white matter would be rejected as inconsistent; the reconstruction would then be the set of sources and source amplitudes that best match the observations and are consistent with the anatomy. This approach has promise for future development.

Such constrained reconstructions have been demonstrated by computer simulation for distributed source models, where they can be easily implemented by excluding elementary sources inconsistent with the anatomy. The particular algorithms used can also incorporate prior information about the expected amplitude of the sources, expected amplitude of the detector noise, and spatial and temporal correlations between sources. The use of such prior knowledge

has been demonstrated in computer simulation to improve the accuracy of the reconstructed source distribution. One potential problem is that true source distributions inconsistent with the assumed prior distributions can generate reconstruction artifacts such that the reconstruction with priors is less accurate than the reconstruction with no priors.

10.5 Summary

As with electroencephalography and electrocardiology, clinical and scientific applications of MSI await the development of less expensive instrumentation as well as practical computational strategies for determining the spatial distribution of the current generators in the brain and heart. The steady improvement in this technology has been mostly through development of larger arrays of SQUID detectors. Refinements in applied mathematical approaches are probably less important than improvements in computational practicality. The latter presents the greatest challenge.

10.6 Research Opportunities

- For magnetic source imaging to become an acceptable procedure, the cost of biomagnetic instrumentation must first be reduced. The most expensive component is the screening against external interference, given that the necessary soft magnetic material requires a costly manufacturing process. Signal processing algorithms to suppress such interference are available and can be improved, but they may not be sufficient to completely obviate the need for shielding.
- Point and distributed source reconstructions are both approximations of the true distribution of current in the body, but it is not known which is a better approximation. The two types of reconstructions are so different that it is difficult to directly compare their accuracy. It would be beneficial to find some effective method of comparison.
- The incorporation of realistic body geometry and conductivity distributions used in the forward models is an important focus for study. As models become more detailed, they also become more expensive to compute, and additional research in the designs of efficient algorithms is likely to be useful. One question of particular interest is how accurate the forward model must be to yield accurate reconstructions from data with some specified noise level; since improved accuracy costs more computer time, it is preferable to use a forward model that is no more accurate than necessary.
- Existing distributed source models can exploit prior information about the expected source amplitudes and both spatial and temporal correlations between sources. It is not known, however, what values should be adopted for these prior distributions to optimize the trade-off between improved accuracy and an increased risk of artifacts. Reconstruction methods using prior information can create reconstruction artifacts when the true source distribution is inconsistent with the assumed priors. Methods are needed for detecting when priors are inconsistent with the data.

10.7 Suggested Reading

1. Ahonen, A.I., Hämmäläinen, M.S., Kajola, M.J., Knuutila, J.E.T., Laine, P.P., Lounasmaa, O.V., Simola, J.T., Tesche, C.D., and Vilkmann, V.A, A 122-channel magnetometer covering the whole head, in *Proceedings of the Satellite Symposium on Neuroscience and Technology of the 14th International Conference of the IEEE Engineering in Medicine and Biology Society*, A. Dittmar and J.C. Froment, eds., Lyon, 1992, 16–20.
2. Bertrand, O., Thevenet, M., and Perrin, F., 3D finite element method in brain electrical activity studies, in *Biomagnetic Localization and 3D Modelling*, J. Nenonen, H.M. Rajala, and T. Katila, eds., Helsinki University of Technology, Department of Physics, Report TKK-F-A689, Helsinki, 1991, 154–171.
3. Cohen, D., Cuffin, B.N., Yunokuchi, K., Maniewski, R., Purcell, C., Cosgrove, G.R., Ives, J., Kennedy, J., and Schomer, D., MEG versus EEG localization test using implanted sources in the human brain, *Ann. Neurol.* **28** (1990), 811–817.
4. Hämmäläinen, M.S., and Sarvas, J., Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data, *IEEE Trans. Biomed. Eng.* **BME-36** (1989), 165–171.

5. Hughett, Paul, An optimal constrained linear inverse method for magnetic source imaging, *Ann. Biomed. Eng.*, accepted for publication.
6. Ioannides, A.A., Bolton, J.P.R., and Clarke, C.J.S., Continuous probabilistic solutions to the biomagnetic inverse problem, *Inverse Problems* **5** (1989), 999–1012.
7. Kaufman, Lloyd, Kaufman, James H., and Wang, Jia-Zhu, On cortical folds and neuromagnetic fields, *Electroencephalogr. Clin. Neurophysiol.* **79** (1991), 211–226.
8. Lütkenhöner, B., On the biomagnetic inverse procedure’s capability of separating two current dipoles with a priori known locations, in *Advances in Biomagnetism: Clinical Aspects, Excerpta Medica*, M. Hoke, S.N. Ern e, Y.C. Okada, and G.L. Romani, eds., Amsterdam, 1992, 687–692.
9. Menninghaus, E., Lütkenhöner, B., Gonzalez, S.L., Hampson, S., and Pantev, C., Localization of a dipolar source in a skull phantom: Comparison between boundary element model and spherical model, in *Proceedings of the Satellite Symposium on Neuroscience and Technology of the 14th International Conference of the IEEE Engineering in Medicine and Biology Society*, A. Dittmar and J.C. Froment, eds., Lyon, 1992, 119–122.
10. Moran, J.E., Jacobson, G.P., and Tepley, N., Finite difference field mapping, in *Advances in Biomagnetism: Clinical Aspects, Excerpta Medica*, M. Hoke, S.N. Ern e, Y.C. Okada, and G.L. Romani, eds., Amsterdam, 1992, 801–805.
11. Nenonen, J., Purcell, C.J., Horacek, B.M., Stroink, G., and Katila, T., Magnetocardiographic functional localization using a current dipole in a realistic torso, *IEEE Trans. Biomed. Eng.* **BME-38** (1991), 658–664.
12. Purcell, Christopher J., and Stroink, Gerhard, Moving dipole inverse solutions using realistic torso models, *IEEE Trans. Biomed. Eng.* **BME-38** (1991), 82–84.
13. Ribary, U., Ioannides, A.A., Singh, K.D., Hasson, R., Bolton, J.P.R., Lado, F., Mogilner, A., and Llinas, R., Magnetic field tomography (MFT) of coherent thalamo-cortical 40-Hz oscillations in humans, *Proc. Natl. Acad. Sci. USA* **88** (1992), 11037–11041.
14. Stroink, Gerhard, Cardiomagnetic imaging, in *Frontiers in Cardiovascular Imaging*, Barry L. Zaret, Leon Kaufman, Alan S. Berson, and Rosalie A. Dunn, eds., 161–177, Raven Press, New York, 1993.
15. Wang, Jia-Zhu, Williamson, Samuel J., and Kaufman, Lloyd, Magnetic source images determined by a lead-field analysis: The unique minimum-norm least-squares estimation, *IEEE Trans. Biomed. Eng.* **BME-39** (1992), 665–675.

Chapter 11

Medical Optical Imaging

11.1 Introduction

The past few years have seen considerable advances in devices and algorithms designed to give useful clinical information through the use of a laser beam as a probe into human tissue. The two main areas of interest are tissue spectroscopy—for establishing the presence and if possible the concentration of certain quantities of interest—and imaging, that is, producing an image of these substances as they are localized in different parts of the tissue. The mathematical model is presented in section 14.1.4.

The term laser optical tomography has sometimes been used, because data are collected by applying a laser source at one or different locations around the object of interest and then detecting the light emitted in one or several locations to determine certain characteristics of the medium transversed by the beam. The method is thus analogous to x-ray computed tomography (CT).

Light in the near-infrared range (wavelength from 700 to 1200 nm) penetrates tissue and interacts with it in complicated ways; the predominant effects are absorption and scattering. Many of the substances of interest, such as hemoglobin and cytochromes, exhibit characteristic absorption spectra that depend on whether the molecule is in its oxidized or reduced state. Other substances of great importance, such as NAD/NADH (nicotinamide adenosine diphosphate), exhibit fluorescence properties that allow for their detection after excitation by light. As these substances play crucial roles in metabolic processes at the cell level, the ability to discern them through indirect measurements would have important medical implications.

An early tool of medical optical imaging was the “oximeter,” devised in the 1930s to detect the amount of oxygen in blood by measuring the ratio of the light absorbed at two wavelengths. Great improvements to this concept came in the 1970s with the advent of microprocessors and light-emitting diodes that permitted the use of many more wavelengths, thus allowing measurement of the absolute amount of oxygen and elimination of background effects. Assessment of the oxygen content of arterial blood through such methods has become a major diagnostic tool for studying acutely ill patients.

The potential of imaging with light was reinforced with the successful application of optical tools to determine the levels of oxygen in the brain of a cat. Later, this concept was used in monitoring brain and muscle oxygenation in humans, as well as in other applications.

Optical imaging devices are now used in the neonatal clinic at the Stanford University Medical Center, the University Hospital in London, and possibly elsewhere, and Hamamatsu Corp., in Japan, produces an infrared spectrometer for bedside monitoring of oxygenation in the brains of babies.

11.2 Data Acquisition Strategies

Objects have been illuminated for medical imaging by the continuous beam, time-resolved beam, and phase-modulated beam methods. The last method has the advantage of allowing measurement of the mean optical pathlength without the size and cost problems associated with ultrashort light pulses and a fast optical detector. The phase-modulated beam as a measure of optical length has proved to be very useful in quantifying absorption and scattering coefficient measurements.

Data collection methods can be divided into two broad classes: (1) those that try to isolate photons undergoing no or very little scattering from source to detector and that thus may be able to use simpler reconstruction algorithms because the path of each photon is considered known; and (2) those that collect light over a longer period, with or without taking into account photon arrival times.

Preliminary mathematical analysis indicates that the preferred approach is to sort photon counts by arrival time and then proceed to solve the complex nonlinear inversion problem.

11.3 Comparisons with Other Imaging Modalities

Comparisons between optical imaging and more established modalities of medical imaging such as CT, positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) are both inevitable and pedagogically useful. It is quite possible that medical optical imaging's eventual role will be as a complement to standard techniques rather than as a replacement for them, or perhaps as a complement to other modalities that are under development and have not yet reached maturity (e.g., electrical impedance tomography or magnetic source imaging).

In a broader sense it appears likely that the effective use of optical imaging technology will require advances in understanding how low-energy photons interact with complex media like human tissue, particularly as the energy of the photons involved is decreased. Increased understanding of energy-tissue interaction could be applied in some of the more established imaging modalities, many of which do not model for scattering events and strive to throw out this “noise.” One result might be the development of good ways of dealing with “scatter correction” in other imaging modalities.

In thinking about the possibilities and difficulties associated with optical imaging, it is necessary to realize that energy in the infrared range interacts with tissue in ways that are dominated by scattering, a phenomenon that is to a large extent absent when higher-energy portions of the electromagnetic spectrum are used. Although the scattering in optical imaging is at times confined to the forward direction, scattering effects accumulate over any thick specimen—because the scattering coefficients in tissue are rather large—and the directional beam of laser light is quickly converted into a diffuse flux. In fact the situation is more complicated: large particles with diameters much larger than that of the wavelength act as mostly forward scatterers, while cell nuclei and mitochondria within the cell scatter light almost uniformly in all directions.

In the case of CT one always assumes that a reading connected to a given source-detector pair reflects an absorption event that took place along the straight line joining the two. Mathematically this assumption allows reduction of the problem to one of reconstructing an unknown attenuation distribution function from its straight line integrals. Nothing like that can be assumed when the photons travel according to the local and varying characteristics of the medium. In fact the photon counts for any source-detector pair result from many absorption and scattering events that happen over the entire medium. It is fortunate that in the near-infrared range, the absorption coefficient of water, fat, and the other chromophores is relatively low.

The physically derived mathematical models of the interaction of light and any complex media, such as human tissue, usually take the form of (1) a Boltzmann-type linear transport equation or (2) some of its approximations, such as the diffusion equation with a diffusion coefficient and an absorption coefficient that model the (unknown) local characteristics of the medium (cf. section 14.1.4). The problem of interest in imaging is then that of determining these functions from the data. A more complete and realistic model allows for an anisotropic diffusivity, which makes the inversion problem even more complex.

The mathematical problem of solving simultaneously for an unknown pair of spatially varying diffusion and absorption coefficients given data collected under a variety of imaging conditions has only recently started to attract attention. More work is required—both to obtain mathematical formulations of complicated scattering environments and to assess the numerical stability of possible inversion algorithms for these highly nonlinear problems. Existing mathematical theory makes it likely that any reconstruction algorithms that emerge could be seriously ill-conditioned or sensitive to noise in the data.

Comparison with the corresponding mathematical problem in CT is likely to lead to pessimism about the prospects for handling the complexities of optical imaging. CT involves a linear inversion problem, for which a well-developed theory was available, although the pioneers of the field had to develop practical implementation schemes. Questions about the degree of ill-conditioning of this problem were answered, and an effective, fast, and accurate way of implementing the inversion was found rather swiftly, so that CT became a great success story. The fact that even the simpler model of optical imaging involves two functions (attenuation and scatter) makes the present problem akin to that encountered in SPECT and PET.

Thus one can ask the question, Is laser optical tomography ever going to yield pictures of the “visual” quality of CT scans? The answer is most likely not. The question then is, What role could this emerging technique play? This issue is addressed below with full knowledge of the fact that, in the past, many such predictions have been incorrect or shortsighted.

11.4 Possible Applications of Optical Tomography

The purpose of any new medical imaging technology is to help the clinician reach a decision. This decision is based on both an assessment of the tissue structure and knowledge about the functioning state of the tissue. There is wide consensus that the emerging technology of optical tomography could be useful in providing both kinds of information. In neonatal imaging, optical tomography has been used to assess tissue structure in ongoing trials to monitor the location and onset of hemorrhage (using both time-resolved and continuous-beam systems); in efforts to monitor strokes, it has been used to gain knowledge about functioning tissue, a task that does not require the high spatial image resolution needed for hemorrhage detection. Developers of optical technology must maintain a clear understanding of which of these two kinds of information is the goal of their particular application.

Applications in which optical tomography in its present state has started to yield useful results, and where high spatial resolution is not required, include the measurement of tissue oxygenation for the study of muscular dystrophy, assessment of tissue perfusion in the extremities of diabetics, and the study of brain activity during specific tasks. Some topics that have been or could be studied as part of this effort include the following:

- Changes in the concentration of glucose modify the indexes of refraction of both the intra- and extracellular medium, which alters the scattering properties of those media. Recently, such changes in scattering have been monitored separately from photon absorption in clinical measurements, and a very good correlation has been observed between a reduced scattering coefficient and the concentration of glucose determined from a blood test, suggesting the possible use of optical methods for glucose monitoring.
- In certain situations it is possible to identify most of the received photons with the blood in the smaller blood vessels. This possibility could be useful since it is only in the capillary beds that oxygen exchange with tissue occurs. Measurement of the levels of hemoglobin saturation in these vessels is important in the monitoring of metabolic function.
- Although blood vessels and capillaries occupy only a very small portion of brain tissue, any hemorrhage results in a relatively high local concentration of hemoglobin. Thus cranial bleeds could be detected by the asymmetry of absorption with a simple device usable in the emergency room or intensive care unit.
- Contrast agents such as indocyanine green, which absorbs at 800 nm and fluoresces at a slightly shorter wavelength, can be detected in very small amounts. This property has been used to observe regions with high differential absorbance in the human breast, suggesting the possibility of optical mammography.
- The fluorescence properties of certain dyes depend on the intensity of the excitation and biological parameters such as oxygenation state and local electric potential. Certain dyes are sensitive to the drop in potential that exists across many cell walls. Optical scattering experiments could help in the study of these potentials across an array of cells. The results could in turn be important in the study of wound healing, fetal development, and other organ processes including brain function.
- A combination of coherent imaging and spectroscopic measurements could have applications in the study and monitoring of thin tissues, such as the skin and many parts of the eye, where the effects of scattering are not severe. The living cornea, lens, and retina could be visualized in detail and their functions studied through careful optical scattering investigations.

11.5 Research Opportunities

The central question in investigations of medical optical imaging is, How do human tissues and other complex media propagate light, and what can be learned about those media through such study? Specific examples of this line of research are given in the preceding section. Important general directions are as follows:

- Development of mathematical tools for the mapping of parameters such as absorption and scattering coefficients that play a crucial role in the propagation of light in human tissues,
- Investigations to determine the biophysical basis for diffusion of light in tissue, and
- Investigations of excited fluorochromes that target specific tissues.

11.6 Suggested Reading

1. Arridge, S.R., Cope, M., and Delpy, D.T., The theoretical basis for the determination of optical pathlengths in tissue—Temporal and frequency analysis, *Phys. Med. Bio.* **37** (1992), 1531–1560.
2. Arridge, S.R., Schweiger, M., Hiraoka, M., and Delpy, D.T., A finite element approach for modeling photon transport in tissue, *Med. Phys.* **20** (1993), 299–309.
3. Benaron, D.A., Optical biopsy and imaging advance medical care, *Laser Focus World* **30** (1994), 79.
4. Benaron, D.A., Benitz, W.E., Ariagno, R.L., and Stevenson, D.K., Noninvasive methods for estimating in vivo oxygenation, *Clin. Pediatr.* **31** (1992), 258–273.
5. Benaron, D.A., and Stevenson, D.K., Optical time-of-flight and absorbance imaging of biologic media, *Science* **259** (1993), 1463–1466.
6. Boas, D.A., O’Leary, M.A., Chance, B., and Yodh, A.G., Scattering and wavelength transduction of diffuse photon density waves, *Phys. Rev. E* **47** (1993), R2999–R3002.
7. Chance, B., Kang, K., He, L., Weng, J., and Sevick, E., Highly sensitive object location in tissue models with linear in-phase and anti-phase multi-element optical arrays in one and two dimensions, *Proc. Natl. Acad. Sci. USA* **90** (1993), 3423–3427.
8. Chance, B., Leigh, J.S., and Gratton, E., New image, *Nature* **349**, (1991), 660.
9. Chance, B., Zhuang, Z., Unah, C., Alter, C., et al., Cognition-activated low-frequency modulation of light absorption in human brain, *Proc. Natl. Acad. Sci. USA* **90** (1993), 3770–3774.
10. Fantini, S., Franceschini, M.A., Fishkin, J.B., Barbieri, B., et al., Quantitative determination of the absorption spectra of chromophores in strongly scattering media—A light-emitting-diode based technique, *Appl. Opt.* **33** (1994), 5204–5213.
11. Fantini, S., Franceschini, M.A., and Gratton, E., Semi-infinite-geometry boundary problem for light migration in highly scattering media—A frequency-domain study in the diffusion approximation, *J. Opt. Soc. Am. B* **11** (1994), 2128–2138.
12. Firbank, M., Hiraoka, M., Essenpreis, M., and Delpy, D.T., Measurement of the optical properties of the skull in the wavelength range 650–690 nm, *Phys. Med. Biol.* **38** (1993), 503–510.
13. Fishkin, J.B., and Gratton, E., Propagation of photon-density waves in strongly scattering media containing an absorbing semi-infinite plane bounded by a straight edge, *J. Opt. Soc. Am. A* **10** (1993), 127–140.
14. Gopinath, S.P., Robertson, C.S., Grossman, R.G., and Chance, B., Near-infrared spectroscopic localization of intracranial hematomas, *J. Neurosurg.* **79** (1993), 43–47.
15. Gratton, G., Maier, J.S., Fabiani, M., Mantulin, W.W., et al., Feasibility of intracranial near-infrared optical scanning, *Psychophysiology* **31** (1994), 211–215.
16. Grünbaum, F.A., and Patch, S.K., Simplification of a general model in diffuse tomography, in *Proc. Photon Migration and Imaging in Random Media and Tissues*, Britton Chance and Robert R. Alfano, eds., *SPIE Proc.* **1888** (1993), 387–401.
17. Haselgrove, J.C., Wang, N.G., and Chance, B., Investigation of the nonlinear aspects of imaging through a highly scattering medium, *Med. Phys.* **19** (1992), 17–23.
18. Hebden, J., Delpy, D., and Arridge, S., Infrared lasers muscle in on medical imaging, *Physics World* **6** (1993), 23–24.
19. Hiraoka, M., Firbank, M., Essenpreis, M., Cope, M., et al., A Monte-Carlo investigation of optical pathlength in inhomogeneous tissue and its application to near-infrared spectroscopy, *Phys. Med. Biol.* **38** (1993), 1859–1876.
20. Kantar, A., Giorgi, P.L., Gratton, E., and Fiorini, R., Probing the interaction of PAF with human platelet membrane using the fluorescent probe laurdan, *Platelets* **5** (1994), 145–148.

21. Klibanov, M.V., Gutman, S., Barbour, R.L., Chang, J., Malinsky, J., and Alfano, R.R., Consideration of solutions to the inverse scattering problem for biomedical applications, in *Proc. Physiological Imaging, Spectroscopy, and Early-Detection Diagnostic Methods*, Randall L. Barbour and Mark J. Carvlin, eds., *SPIE Proc.* **1887** (1993), 77–97.
22. Kurth, C.D., Steven, J.M., Benaron, D., and Chance, B., Near-infrared monitoring of the cerebral circulation, *J. Clin. Monitoring* **9** (1993), 163–170.
23. O’Leary, M.A., Boas, D.A., Chance, B., and Yodh, A.G., Reradiation and images of diffuse photon density waves using fluorescent inhomogeneities, *J. Lumin.* **60** (1994), 281–286.
24. Peebles, D.M., Edwards, A.D., Wyatt, J.S., Bishop, A.P., et al., Changes in human fetal cerebral hemoglobin concentration and oxygenation during labor measured by near-infrared spectroscopy, *Am. J. Obst. Gynecol.* **166** (1992), 1369–1373.
25. Schweiger, M., Arridge, S.R., and Delpy, D.T., Application of the finite element method for the forward and inverse problems in optical tomography, *J. Mathematical Imaging and Vision* **3** (1993), 263–283.
26. Sevick, E.M., Chance, B., Leigh, J., Nioka, S., et al., Quantization of time-resolved and frequency-resolved optical spectra for the determination of tissue oxygenation, *Anal. Biochem.* **195** (1991), 330–351.
27. Singer, J., Grünbaum, F.A., Kohn, P.D., and Zubelli, J.P., Image reconstruction of the interior of bodies that diffuse radiation, *Science* **248** (1990), 990–993.
28. Wang, L., and Jacques, S.L., Hybrid model of Monte-Carlo simulation and diffusion theory for light reflectance in turbid media, *J. Opt. Soc. Am.* **10** (1993), 1746–1752.
29. Wilson, B.C., Sevick, E.M., Patterson, M.S., and Chance, B., Time-dependent optical spectroscopy and imaging for biomedical applications, *Proc. IEEE* **80** (1993), 918–930.

Chapter 12

Image-Guided Minimally Invasive Diagnostic and Therapeutic Interventional Procedures

Medical imaging is applied at various stages in the patient management process: screening, diagnosis, staging, therapy delivery, and therapy follow-up. The primary role for imaging has been diagnostic, but there is increasing, albeit limited, use of medical imaging modalities—including endoscopy, x-ray fluoroscopy and computed tomography (CT), ultrasound, and recently magnetic resonance imaging (MRI)—for interventional diagnostic and therapeutic uses, driven by the thrust for minimally invasive procedures. Medical imaging permits the physician to plan the therapeutic procedure more accurately before carrying it out, to guide the intervention and more correctly locate the position of the interventional tool with respect to the anatomy, to monitor the intervention as it is being carried out, and to control the intervention for optimum results.¹

Although static images are sufficient for planning radiation therapy or providing some anatomical information for surgery, real-time image guidance has been fundamental to the evolution of interventional radiological procedures. More recently, the emergence of minimally invasive therapeutic procedures has encouraged the tendency for surgical procedures to shift from large explorations toward approaches with limited access and restricted visibility. These new applications increase the demand for image guidance and compel the use of the most advanced imaging methods. It is expected that allowing medical imaging systems to play a more direct role in interventional and therapeutic procedures—thus enabling greater precision, increasing foreknowledge, and facilitating even less invasive surgical access—will reduce the cost of patient management and improve the quality of patient outcomes.

Image-guided therapy is a new, emerging field that has close relationships to interventional radiology, minimally invasive surgery, computer-assisted visualization, and robot-assisted surgery.

12.1 Therapeutic Intervention Experience with Different Imaging Modalities

12.1.1 X-Ray Imaging

Static x-ray imaging has been used to allow the therapist to plan radiation therapy and, to a lesser extent, surgery; however, the projection nature of the image limits its value, particularly when tomographic images can be available for this purpose. For guidance, control, and monitoring operations during a procedure, fluoroscopic images are used rather than static images. Fluoroscopy has the advantage of presenting real-time images. Although it is not tomographic, overlying structures are eliminated by using contrast media to accentuate the specific anatomy (usually blood vessels) to be studied. A major disadvantage of fluoroscopic methods is that the x-ray dose delivered to the patient and to the interventionalist can be high. However, x-ray fluoroscopy is the key imaging technique used in interventional procedures.

¹ NOTE: This chapter is adapted in large part from Jolesz and Blumenfeld, 1994; see suggested reading list at end.

12.1.2 Computed Tomography

Computed tomography (CT) imaging has been used in interventional applications in basically two ways: (1) to provide information for planning radiation therapy and for surgery—including in particular stereotactic biopsy, stereotactic craniotomy, and modern stereotactic radiosurgery—as well as for planning conventional radiation therapy; and (2) to provide guidance for image-guided biopsy of various body parts.

The advantage of CT over projection x-ray imaging is that CT is tomographic, presenting the anatomy on a slice-by-slice basis for more exact localization. It is not preferred, at this time, over standard angiographic methods for vascular interventions, because of the difficulty of matching the passage of the bolus through the vascular tree with the acquisition of the appropriate slices. However, there are some hints that this matching may be possible and useful using spiral scanning CT. The major disadvantages of CT are that (1) like other x-ray imaging, it uses ionizing radiation, which poses a risk, particularly to the physician, and (2) it is not sensitive to parameters other than electron density and so does not provide the wealth of tissue parameter information that MRI and potentially ultrasound give. Thus CT is sensitive primarily to anatomic rather than physiological changes, although the anatomic changes it detects can also reflect physiological changes. For example, CT can be used to obtain information about abnormal myocardial wall motion through a gated study or electron beam CT. However, it is not sensitive to changes in temperature, diffusion coefficient, or perfusion in the way that MRI is, and so may not be able to provide warning of a change early enough to address the condition while it is still reversible. In addition, CT in general does not provide as good information about the lesion boundaries or margins as does MRI.

CT has much better geometric accuracy than does MRI, because of the intrinsic curved nature of the magnetic field. Thus CT may be preferred in applications for which geometric accuracy is important, such as surgical procedures that rely on prospectively derived images for navigational guidance.

The most widely used interventional CT procedure is for diagnostic biopsy, which has been applied to the head, neck, thorax, liver, pancreas, adrenal glands, kidney, pelvis, retroperitoneum, and skeletal bone. In addition, CT has been used (1) to guide other percutaneous procedures,² including discectomies, denervations, and neurolysis to relieve pain in the spinal column; (2) via neurolytic techniques, to manage pain due to cancer; (3) to manage fluid collection in the urinary system; and (4) to drain abscesses and other fluid accumulations. These techniques have been described in the book *Interventional Computed Tomography* listed in the suggested reading.

12.1.3 Ultrasound

Ultrasound can be and has been used extensively in interventional radiology, particularly for the guidance of biopsy and to aid in fluid management. The major advantages of ultrasound are its real-time nature and its low cost compared to other imaging modalities. The images are tomographic but are not as clear and crisp as those obtained with CT and MRI, and thus are less acceptable to surgeons.

12.1.4 Endoscopy

The advent of video endoscopic techniques³ revolutionized the field of surgery and made possible the concept of minimally invasive surgery. Originally developed as diagnostic tools, endoscopes and their allied instruments such as laparoscopes, colonoscopes, and bronchoscopes permit the physician to look into the channels of various anatomic tubes (e.g., the gastrointestinal (GI) tract, the genito-urinary (GU) system, the bronchial system) and, when used percutaneously with appropriate measures such as insufflation, into body cavities such as the abdomen. These tools provide high resolution and natural color but, unlike the diagnostic imaging modalities, they provide surface visualization only and cannot indicate how deep a lesion extends below the surface of the organ, nor exactly where in the lumen the field of view is located in relation to the rest of the anatomy.

The marriage of endoscopic surgery with on-line MR guidance promises to provide a significant improvement in the ability of the surgeon or interventional radiologist to understand the operating field and its anatomic context.

12.1.5 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) has specific advantages, over and above its lack of ionizing radiation, for guiding, monitoring, and controlling minimally invasive diagnostic and therapeutic interventions, including the following:

- Superb tissue discrimination, enabling excellent discrimination between various organs, including blood vessels, nerves, and solid organs—without administration of intravenous contrast agents;

²Those performed through the skin, as with a needle.

³Endoscopic techniques use fiber optics to examine the interior of a hollow organ, such as the bladder or stomach.

- Superior definition of the number and extent of lesions and discrimination between the lesions and normal tissue for optimal definition of the target for therapy;
- Best indication of anatomic context of surrounding blood vessels and nerves to provide accurate trajectory definition;
- Spatial resolution that extends into the submillimeter range;
- Direct acquisition of multiplanar and three-dimensional volume images;
- Image updates in well under 1 second with the new rapid image acquisition protocols, including echo-planar imaging (EPI), fast spin echo imaging, and gradient echo imaging, and various hybrid pulse sequences, enabling a close-to-real-time or real-time viewing of physiological motions and the changes induced by the interventional procedures; and
- Ability to characterize functional and physiological parameters of the treated tissues, such as diffusion, perfusion, flow, and temperature.

As a result of these characteristics, MRI can depict the tissue damage induced by various tissue ablative methods⁴ and therefore has the unparalleled potential to not only monitor but also control interventional or minimally invasive surgical procedures.

MRI has potential appeal for most of the minimal access therapy approaches. Intravascular procedures can be significantly improved by imaging vessel walls and the associated pathologies with endoluminal coils. Vascular and endoluminal catheterization can be effectively guided within MRI systems using special localizing coils attached to the catheters.

Angioplasty and atherectomy devices (such as lasers) for vascular applications can be applied more effectively and more safely with on-line imaging control. Laparoscopic and endoscopic vision is limited to a relatively small surface region. This restricted direct visibility can be complemented by obtaining a larger volume to visualize beyond the operational field and below the surfaces without increasing the risk posed by the procedure.

An obvious role of MRI is in monitoring and controlling a variety of methods of interstitial ablative methods, including thermal therapy. In these applications no direct visual inspection is possible since the minimally invasive probe (needle, optical fiber, cryoprobe, RF needles, and so on) is introduced percutaneously or applied by a non-invasive, non-incisional method (e.g., focused ultrasound).

The major disadvantages of MRI in this application are its relatively high cost compared to other imaging modalities, the special environment it creates and requires because of its magnetic field and RF properties, and patient inaccessibility during an examination. In particular, the tools, instruments, and accessories used in the procedure room must be MR compatible and in many, but not all, cases this may require a redesign of the equipment.

12.2 The Roles of Imaging in Therapy

Underlying the use of imaging in therapy is a simple scientific concept that can be implemented at different levels of complexity using various imaging modalities from x-ray radiography to ultrasound, CT, or MRI. This concept is that information obtained from imaging devices can be used to construct “models” of a patient’s particular anatomy that can then be compared with models of normal anatomy and with the executional model in which the actual procedure and its components are described. These models are then projected back to the patient and registered within a single reference frame. The resulting integrated and superimposed models assist the physician who performs an interventional procedure.

In providing better information to the physician to make interventions more effective, imaging can contribute in three key areas—planning, guidance, and monitoring (including localization). In addition, MRI specifically provides image data, which can be used to control the interventional procedure.

12.2.1 Planning

Previously acquired CT and MR images (i.e., those not obtained at the time of a procedure) are being used increasingly for radiation therapy and for surgical planning. The requirements for planning differ somewhat from those for diagnostic imaging. The image sources should be high-quality, high-resolution tomographic images. If sufficient thin-slice contiguous image planes are acquired, they can then be reformatted into other planes or synthesized into three-dimensional renderings of the anatomic information in the images. The different tissues and organs within the images have been previously filtered or segmented appropriately to delineate the target and surrounding structures

⁴ Ablative methods include a variety of means of excising or destroying interstitial tissue.

prior to the creation of the three-dimensional volume and surface renderings so that organs or tissues can be better distinguished and analyzed. At this point appropriate analytic and synthetic operations of trajectory simulation or planning, of probe guidance, and of surgical planning are carried out on the three-dimensional images thus created. Images from various sources, particularly MRI or CT volume rendered images, can be put into register with each other and with real-time video of the patient, specifically for use in guidance during the image-directed procedure.

With the three-dimensional volume rendered image in hand, the physician is now able to create an appropriate surgical plan. Such a plan might delineate the safest and most appropriate route for gaining access to diseased tissue in an organ. It might involve the creation of an appropriate normal template for correcting an anatomic abnormality. Or it might be a suitable road map or series of road maps to guide the repair of a specific organ or anatomic region. Clearly the specifics of the plan will depend on the particular intervention under consideration, for example, the guidance of a biopsy needle to a diseased organ or tissue target or the delivery of a therapeutic dose of a pharmacological agent or heat energy. A range of these interventions is examined below.

12.2.2 Guidance

Once the surgical plan has been created, the next step is to use it to guide the surgery. The interventional or therapeutic tool (a biopsy needle in the simplest case) can be guided in two ways, depending on the imaging capabilities. The preferred method is on-line guidance with rapid update of the images during the procedure to accurately monitor the position of a probe within the anatomy. Fluoroscopy- or ultrasound-guided biopsies are performed routinely and safely. However, in some cases only MRI can provide appropriate target definition and/or anatomical detail for defining the correct trajectory. Real-time MRI-guided biopsies are only now becoming clinically feasible with the advent of MRI systems with appropriately open magnets capable of rapid image update.

Coordinates and trajectories for biopsies or other interventional procedures (such as abscess drainage and nephrostomy) can be obtained from previously derived images. This information remains valid if the patient does not move between the time of imaging and the procedure. Using fiducial marks and measurements,⁵ a “free hand” biopsy can be performed with the guidance of a cross-sectional imaging system. This category of MRI-guided biopsies has been used clinically with the aid of MRI compatible needles and appropriately visible markers.

“Dead-reckoning” navigational methods are more feasible for precise biopsies and for minimally invasive surgical procedures. The stereotactic techniques employ acquired images with stereotactic frames or without them (frameless stereotaxy). In neurosurgery, various rigid frames are fixed to the skull to facilitate transfer of the correct coordinates from one reference system to another. Using this system, imaging data (plain film, CT, or MR) can be transferred to the operating room when the actual procedure is executed. New developments attempt to eliminate the frame while keeping track of the patient’s position and the interventional tool, and relating these simply, flexibly, and relatively inexpensively to the image coordinate system. The tools (e.g., biopsy needles) can be attached to a mechanical arm with optical or electromagnetic sensors that track the movement of the tool. The image data are then registered to the patient’s anatomy so that trajectories for biopsies can be defined from images or, conversely, so that the position of the interventional tool within the patient can be depicted on the displayed images. While both types of stereotactic techniques are used in neurosurgery, only frameless stereotactic methods are feasible for other parts of the body.

There are three important requirements for a guidance system to ensure that it can be used appropriately in a wide variety of interventional situations.

- **Immobilization.** It is often necessary to immobilize the anatomy under consideration, especially when dead-reckoning navigational guidance is used, so that the position of the target remains invariant with respect to the fiducials between the imaging process and the subsequent interventional process. The immobilization tools are different for different parts of the body. They are straightforward for the head, less so but still simple for the limbs, and more difficult for the breast. Targets such as the liver and other internal organs affected by respiratory and cardiac motion can be only partially immobilized, and further development of appropriate techniques may be necessary.
- **Registration.** Registration permits fusion of images in such a way that the coordinate systems of the two images are coincident. A preferred approach would eliminate the frame or other obtrusive fiducial system. Frameless stereotaxy, a proprietary method of registration without frames, permits registration of the patient coordinate system with the three-dimensional patient model or the surgical planning model. This method utilizes a video image or a laser-scanner defined surface area of the patient made during the procedure with the three-dimensional model created in the planning phase from the prospectively acquired images. Such registration techniques not only are useful in the operating theater, but also can be used preoperatively in the planning stage to allow the physician to draw the surgical plan directly on the patient’s anatomy for

⁵That is, those regarded as or employed as a standard of reference, as in surveying.

accurate location of the entry point and projected positions of key structures. Video registration with the three-dimensional model has been used successfully in neurosurgery and can be extended to other parts of the body.

- **Tracking.** A tracking method specifies the position of surgical tools or other instruments during the operation with respect to the image and patient coordinate systems, and also can be used to follow the motion of the patient's body parts, all concurrently. Probably the most versatile tracking system is optical, although the most widely used system is the ISG wand, which uses a mechanical "arm" with encoders on the joints to keep track of the position of the tip. A prototype optical system using the Pixsys FlashpointTM system is being developed for a vertical gap open interventional MRI system and could be adapted to other modalities. This system consists basically of light-emitting diodes mounted on instruments whose positions can be determined accurately by triangulation via a number of line scan video cameras mounted in the room.

12.2.3 Monitoring and Localization

On-line, real-time imaging of the delivery of therapy or of an interventional procedure allows the physician to watch his or her progress. X-ray fluoroscopy has been the modality most used for vascular interventions and for cardiac catheterizations. For needle biopsies and some other interventional procedures, real-time ultrasound represents a viable alternative.

It is only now, with MRI technology being able to acquire the data to create images in less than 1 second, that MRI can be used to monitor and localize probes on-line within the target. In addition, an open magnet configuration is necessary to provide the physician with the access to interact with the patient during an operation. Such magnet configurations with sufficient field strength, appropriate imaging pulse sequences, and vertical gaps affording direct access by the physician to the patient are now becoming available for clinical evaluations.

12.2.4 Control

Immediate visualization of the physiological effect of an intervention, such as increased blood flow, temperature change, or perfusion change, provides the physician the potential for sophisticated control of the procedure. This is of particular importance when the physiological effect is a temperature change and care must be taken to ensure that the temperature change is restricted to a specific target volume and is not too great in any specific anatomic area, and that the spatial localization of the temperature change is appropriate.

Conventional surgery uses direct visual control and eye-hand coordination, but this approach has significant drawbacks. Dissection exposes surfaces only, and so the surgeon cannot see effects below the surface and has to approach the target, dissecting carefully layer by layer. This limitation of direct visual control confined the use of surgical lasers to relatively low penetration and doomed the use of interstitial laser therapy, cryosurgery, and focused ultrasound surgery until appropriate imaging methods, combined with the capability to control energy deposition, were developed.

Direct visualization by the naked eye as is done with endoscopic, laparoscopic, or microscopic techniques utilizes color, texture, and other visual clues to attempt to distinguish normal from abnormal tissues. However, these indicators are often insufficient for accurate characterization. Nor can such modalities permit visualization of the extent of a tumor below the exposed surface or allow the effects of interventions below the surface to be followed.

As detailed above, subsurface visualization is a particular strength of MRI. Specifically, MRI, with pulse sequences that are quite sensitive to temperature changes, can provide real-time, three-dimensional, high-resolution maps of temperature changes during energy deposition or ablation and may be able to signal when tissue proteins are denatured (coagulation) or frozen by phase changes in the image. Most important, the temperature sensitivity of MRI is sufficient to show thermal changes before irreversible tissue damage occurs. This capability allows the operator to modify the amount or rate of deposited energy in time to restrict irreversible changes in the target volume. The detection of temperature changes below the threshold of tissue damage is a great advantage of MRI over ultrasound monitoring of thermal therapy. Ultrasound monitoring is based either on the creation of "bubbles" during tissue coagulation (an adverse result for many thermal therapy procedures) or on the distinction between normal and irreversibly coagulated tissue, whereas MRI can distinguish heated tissue volumes from coagulated ones.

12.3 Thermal Surgery

Image-guided interstitial thermal therapy or surgery is the process of optimally defining a target volume using diagnostic imaging techniques and then destroying all the tissue cells within the target volume by inducing a localized temperature change in the target volume only. Thermal therapeutic techniques use heat or freezing (cryosurgery) to induce irreversible tissue ablation by denaturing the proteins or by destroying cells with ice crystals. For control

of the energy deposition, either multiple temperature probes or specialized temperature-sensitive imaging methods (particularly MRI) are employed.

Localized thermal therapy or “heat surgery” is different from classical hyperthermia treatment. Conventional hyperthermia treatment induces a relatively small temperature elevation (to $\sim 41\text{--}42^\circ\text{C}$) within relatively large tissue volumes in order to exploit the differential sensitivity of neoplastic cells over normal tissue via an otherwise not well understood heat-related damage mechanism. MRI guidance for conventional hyperthermia treatment is feasible and has been tested clinically.

The physical and biological consequences of localized high temperature “thermal therapy” have been experienced by all who have inadvertently burned themselves and are well understood, clear, and not controversial. In the ideal thermal therapy procedure the targeted tissue volume is heated beyond $57\text{--}60^\circ\text{C}$, which is the threshold for protein denaturing. This treatment results in irreversible cell damage in both normal and neoplastic tissues. For this reason, localized thermal therapy is more comparable to surgery than to hyperthermia treatment. The intent is to maintain a tight target volume during the procedure. The applied temperature therefore should be localized exclusively on the target, should not exceed 100°C , and should be applied in relatively short energy pulses to reduce thermal diffusion and minimize heat removal by blood flow and/or perfusion.

Thermal surgery has not been used extensively in the past because of the lack of good imaging guidance and monitoring and because there were not ways to create on-line, high-resolution, three-dimensional temperature maps of the volume of interest for monitoring or control. MRI is sensitive to temperature changes through the temperature sensitivity of the T_1 relaxation time, the diffusion coefficient, or the chemical shift parameters, and recent advances have made it possible to obtain MR images in less than 1 second, thus making it feasible to obtain and update three-dimensional temperature change “maps” of the tissue under consideration in times matched to the temporal resolution of the thermal changes so as to avoid artifacts. This rapidity allows MRI to be used to guide, monitor, and control the heat surgery.

12.3.1 Interstitial Laser Therapy

Delivery of thermal therapy can be percutaneous or non-incisional. Percutaneous delivery via optical fibers coupled to lasers is a promising application for tumors accessible via needle insertion and has been widely used and reported on. MRI monitoring of interstitial laser therapy has been suggested, and animal studies demonstrating its feasibility have been carried out. Initial clinical applications for interstitial laser therapy include stereotactic brain tumor surgery and liver surgery, and MRI localization and guidance without imaging during laser treatment have been demonstrated in a head-neck procedure. However, laser-induced thermal surgery has the drawback of a relatively high temperature gradient at the tip of the optical fiber. The tissue at the tip is heated well above 100°C , creating vapor and “smoke” and, more importantly, a non-uniform thermal lesion through the treated volume. Furthermore, the treatable volume is limited, and it is not normally matched to the target shape without moving the fiber and using multiple probes or multiple needle penetrations. Computerized control of this procedure has been suggested but has not yet been implemented because significant progress must be made in understanding the structure and biology of the target tissues, as well as their optical and thermal properties and how they change during irradiation, before laser-induced thermal surgery can be fully automated.

12.3.2 Cryotherapy

Freezing causes cell destruction through the development of ice crystals, which gives cryotherapy some advantages over heat surgery because the collagen structure of the tissues is not destroyed, vessel walls are preserved, and reinnervation is possible. Current cryosurgical procedures are open (liver) or percutaneous (prostate) and require imaging guidance and monitoring. Ultrasound is used to indicate the expanding freezing zone but can show only the proximal freezing boundary, so that the shape of the remote boundary must be inferred from symmetry assumptions. These assumptions are not very accurate, particularly in patients with fibrotic tissue, such as those who have undergone radiation therapy.

Recent technological developments in the cryoprobes, which have reduced their diameter and made them MRI compatible, have made it possible to monitor cryotherapy by MRI. The lack of signal within the frozen zone sharply delineates the treated area, and the expansion of the freezing zone is well manifested on “real-time” MRI. The open magnet configuration should make cryosurgery under MRI control feasible for a wide variety of new procedures.

12.3.3 Focused Ultrasound

Focused ultrasound (FUS) evolved from lithotripsy and has affinities to localized hyperthermia techniques pioneered in the 1940s and 1950s. FUS uses an ultrasound transducer to create a point source of heat at its focus. High-aperture ultrasound transducers can create a converging beam focused to a high-intensity zone. Within this focal

volume various quantities of thermal energy can be deposited without any obvious damage to the surrounding tissue. The point source of heat can be made sufficiently hot to denature protein in a cigar-like cylindrical shape varying between approximately 1–3 mm in diameter and 2–5 mm in length by “instantaneously” (under 1 s) raising the temperature to between 70°C and 100°C. In order to contain the heat in a well-defined target zone and reduce the effect of thermal diffusion, the heat is delivered in short high-energy pulses. This method results in a sharp temperature drop-off to normal body temperature outside the focal spot and so improves the control of temperature. Blood flow in the target vicinity plays a bigger role in carrying away the heat than does diffusion through the surrounding tissue.

However, experimentation in appropriate tissue and animal models will be required to better understand the most appropriate pulse amplitude and timing protocol. A key determinant will be the issue of cavitation. A high-intensity short-duration pulse may be more likely to create cavitation. Longer-duration, lower-amplitude pulses may also reduce the effect of surface heating, which may become one of the limitations in FUS.

Although this technology has been available for decades, the lack of appropriate localization and temperature-monitoring techniques has made it difficult to achieve clinically useful applications. Currently, ultrasound monitoring is used to localize and target the focused beam and to detect thermally induced irreversible tissue damage.

MRI has been suggested as the optimal technique for spatially localizing, targeting, and controlling heat deposition and has been tested in animal experiments using MRI-compatible transducers positioned within the MRI patient table. The focus can be translated in three dimensions so as to enable the point source of heat to be moved through the target zone. Unlike percutaneous methods, MRI allows scanning of an extended non-spherical target without resorting to multiple punctures or an array of needles. Scanning of the focal spot of heat is computer controlled by “operating” on the MR image of the region under consideration in the patient. Although there are different ways of implementing the scan program, the basic idea is to move, on the MR image, a computer cursor that controls the positioning of the point source of heat through the target region in a sequenced program that ensures that only the desired volume is totally ablated. A test pulse of heat can first be introduced to ensure that the heat is deposited in the appropriate place in the target. This initial pulse of heat or “tracer shot” can be low enough that its effect is reversible without any permanent damage to the tissue. Upon assurance of proper targeting, the system can be switched to therapeutic mode and the target zone scanned. The deposition of heat is monitored by rapid MRI, and the effects are immediately visible on the image. This feedback enables control of the process to avoid damage to normal structures.

The key benefit of FUS is that it is non-incisional and as such has many of the characteristics of an ideal surgical tool (see Table 12.1). It destroys the target and only the target with no entry track of injury. The results are reproducible and predictable because the process can be very accurately controlled. The method results in nonhemorrhagic, sharply demarcated lesions. The surgical effect is localized, is instantaneously visible in the MR image, has no delayed or systemic effects, and is repeatable without additional risk to the patient. Thus, in addition to enabling improved outcomes, MRI-guided FUS should reduce recuperation time, infections, and morbidity.

Because ultrasound beams are blocked by air and bone, treatment of different anatomic locations will likely require different transducer shapes and application methods to optimize the treatment. Ultrasound transducers for neurosurgical procedures or for use with the thyroid will differ from those for the breast or lower abdomen. Treating the upper lobe of a liver may require a flexible transducer head that can be positioned beneath the rib cage pointing cranially. Likewise, transducers may have to be placed in multiple positions on a patient, and hence an MRI system that provides more open access and greater flexibility will be needed. Treatment of the prostate, for example, may require two transducers, one placed externally and one inserted in the rectum, somewhat analogous to the use of the combination phased array coil with insertable rectal coil. The requirements of flexible placement of transducers and simultaneous application of MRI and FUS pose significant design challenges. Realizing the full potential of FUS will require an open MRI system that permits the expanded flexibility required to perform a wide variety of procedures.

FUS systems are under clinical investigation primarily for use with benign prostatic disease and in the kidney and bladder using ultrasound guidance. An MRI-guided FUS system for treating breast tumors has been developed for clinical testing. The clinical potential of MRI-guided FUS is so significant that rapid development of the technique should be encouraged.

12.4 Research and Development Opportunities

Planning

- Definition of tumor and other surgical target margins or boundaries utilizing various medical imaging techniques (correlating with spatially registered histology to estimate the capabilities of the various imaging modalities in defining the boundaries for various anatomic regions).
- Development of real-time image-processing techniques, particularly rapid methods of model creation, three-dimensional rendering, and accurate segmentation of anatomic tissues for various imaging modalities.

Table 12.1. Comparison of Focused Ultrasound with an Ideal Surgical Tool

Ideal Surgical Tool Requirements	MRI-Guided Focused Ultrasound
Should destroy target and nothing but the target	Can destroy preselected targets located deep within tissue without damage to surrounding tissue
No entry track of destruction or injury	Tissue in the entry path is not irreversibly heated
Resultant lesions reproducible and predictable	Lesion size and shape can be accurately predicted and monitored in real time
Size, shape, and position of resultant lesions are finely controllable	Size, shape, and position of lesions in target organ are controllable with MRI guidance
Resultant lesions should be non-hemorrhagic	Nonhemorrhagic lesions produced, because cavitation easier to avoid
Resultant lesions should be sharply demarcated from unaffected surrounding tissue	Margins of the necrotized tissue are sharply demarcated from the surrounding normal tissue and visible in MR images
Resultant lesions should be instantaneous in development	Lesions develop within fractions of a second
No delayed effects such as those produced by radiation therapy	No delayed effects
Should allow creation of transient alterations of function sublethally to assist in target localization	Reversible heat pulse visible on MR image to enable locating the subsequent irreversible pulses
Should enable process to be repeatable without additional risk to the patient	Repeatable without a dose limit (unlike radiosurgery)
Should not require deep anesthesia	Will not require any deep anesthesia
Should be able to reach any part of the body	Blocked by bone and air

NOTE: Column 1 is adapted from P.P. Lele, Ultrasound in surgery, in *Fundamental and Applied Aspects of Non-Ionizing Radiation*, S.M. Michaelson et al., eds., Plenum Press, New York, 1975, 325–340.

- Research in the area of surgical planning and simulation, particularly trajectory planning for needle biopsy, its basic surgical application today.
- Improvement, via more complex automated technologies, of current registration or image fusion methods of different medical imaging modalities and more particularly of video-based and laser-scanning techniques with prospectively created models.

Guidance and Localization

- Development of flexible and untethered sensors to provide anatomic fiducial marks or information on the position of needles, catheters, and surgical instruments, for tracking of instruments or for fusing patient and image coordinate systems.
- Development of computational systems and algorithms to enable “instantaneous” reconstruction, reformation, and display of the image data so as to enable real-time following of a physician’s actions during a procedure (e.g., advancing a catheter or needle).
- Development of methods to update prospectively created models using real-time imaging information acquired during the interventional procedure and reflecting changes effected by the intervention. Real-time image processing and therefore high-performance computing are necessary not only to guide tools to the target but also to monitor and display the changes occurring within image volumes.
- Development of various display technologies, particularly virtual reality and three-dimensional or stereoscopic displays, to provide appropriate interfaces that enable the physician to make the best use of the data. It is expected that these technologies will advance rapidly over the next few years, but driven by other commercial products, and so they will have to be adapted to provide appropriate temporal and spatial resolution suitable for surgical and interventional applications. Such developments are required for monitoring and control tasks as well.

Monitoring

- Definition of the temporal resolution required for various image-guided therapeutic procedures, taking into consideration the physical characteristics of the specific imaging modalities and the dynamic properties of the monitored procedures, specifically for multislice volumetric monitoring.
- For MRI, development of new pulse sequences specifically for therapy applications rather than diagnostic applications. A particularly important need is the development of highly temperature-sensitive pulse sequences to enable monitoring of “heat surgery.”

Control

- Investigations to correlate the factors affecting energy deposition or abstraction (e.g., pulse duration, pulse energy, and power spectrum) with histological and physiological changes in the tissue and resulting image changes, for the purpose of determining mechanisms of thermal damage and the biophysical changes that take place during various thermal surgical procedures such as interstitial laser therapy, cryoablation, and high-intensity focused ultrasound treatment. Such investigations need to be done for various anatomic regions and medical conditions for which such therapy might be appropriate.
- Investigation of the range of medical conditions amenable to treatment with minimally invasive techniques made possible by expanded capabilities for visualization during a procedure via the various medical imaging modalities.

Instruments and Systems

- Although prototypical MRI systems have been created that provide direct and easy access to the patient, more research and development is required to optimize further the geometric configuration of these systems. Similar requirements are appropriate for the other imaging modalities, particularly CT.
- Development of less expensive two-dimensional detector arrays for CT and other x-ray imaging modalities, and of less expensive two-dimensional transducer arrays for ultrasound, along with appropriate means for acquiring, reconstructing, and displaying the data.
- Improved methods of inexpensively shielding the magnetic field to enable inexpensive retrofitting of existing MRI systems into current operating rooms.
- Integration of imaging methods with therapeutic procedures, including feedback systems between data display devices and image information, computer-assisted image-controlled surgical tools, robotic arms, and instruments.
- Creation and development of new instruments and tools to accomplish new tasks enabled by the availability of image-guided therapy. Compatibility of various interventional and surgical tools and imaging systems is important, especially in the case of MRI-guided therapy.

12.5 Suggested Reading

1. Cope, C., ed., *Current Techniques in Interventional Radiology*, 2nd Edition, Current Medicine, Philadelphia, 1995.
2. Jolesz, F.A., and Blumenfeld, S.M., Interventional use of magnetic resonance imaging, *Magn. Reson. Q.* **10** (1994), 85–96.
3. Kadir, S., *Current Practice of Interventional Radiology*, B.C. Decker, Inc., Philadelphia, 1991.
4. Kandarpa, K., and Aruny, J.E., eds., *Handbook of Interventional Radiologic Procedures*, 2nd Edition, Little Brown, Boston, 1995.
5. Kim, D.S., in *Peripheral Vascular Interventions*, Michael D. Darcy, ed., Society of Cardiovascular and Interventional Radiology, Reston, Va., 1994.
6. Seibel, R.M.M., and Gronemeyer, D.H.W., eds., *Interventional Computed Tomography*, Blackwell Scientific Publications, Boston, 1990.

Chapter 13

Frontiers of Image Processing for Medicine

Image processing is based on two major categories of manipulation of arrays of two-dimensional data. The first category includes the restoration of one or more objects by compensating for noise, motion, shading, geometric distortions, and other sources of noise degradation associated with the image acquisition system. The second category involves the enhancement of information and data reduction to emphasize, measure, or visualize important features of the image. In recent years, the field of medical imaging has required that the role of image processing expand from the analysis of individual two-dimensional images to the extraction of information from three-dimensional images, multimodality images, and time sequences of three-dimensional images. This change has prompted the development of sophisticated algorithms for interpretation of multi-dimensional data, a task that is far from accomplished, while placing a strong emphasis on processing speed and available memory for the computer systems performing the analysis.

In the traditional approach, data are analyzed by visual evaluation of cross-sectional slices as represented by gray-scale images on radiological film. The data are typically acquired and filmed by technologists using predefined imaging protocols, and the results are read by radiologists. The films are also used to communicate with referring physicians and for reference during patient procedures. Despite a recent trend toward its use, digital representation of medical images has not yet become generally accepted, and its widespread application will require further developments in the hardware and software for manipulation of image data.

The major difficulty in interpreting cross-sectional gray-scale images is that anatomic structures look very different from their three-dimensional appearance. This discrepancy requires the physician to perform a significant mental translation of the data, a task that requires highly specialized training. Although radiologists undergo such training, the visual interpretation of the data sets becomes observer dependent, and others may have more difficulty in visualizing the data. In view of the relatively large size of a typical three-dimensional data set (e.g., $80 \times 256 \times 256$) and the fact that a single imaging examination may include the acquisition of several such data sets, the radiologist can work more efficiently if the information from many slices is concentrated into one rendering. A composite image also facilitates communication with other clinicians and leads to the possibility of generating quantitative rather than qualitative information from the images.

These observations make it clear that for medical image analysis, the fundamental mathematical need is the derivation of procedures for extracting the clinically important features from one or more large data sets—for example, quantitative information on tumor volume for each of several studies over a time period, to help gauge the efficacy of different treatments, or a parametric map created to represent rate constants from a time series of tracer movements in the brain or heart. The procedures associated with this type of contemporary image analysis can be separated into several different classes:

- Image segmentation,
- Computational anatomy,
- Registration of multimodality images,
- Synthesis of parametric images,
- Data visualization, and
- Treatment planning.

The following discussion presents some of the key mathematical methods being considered for addressing these requirements. See also the related discussions earlier in this report in sections 3.5, 4.3, and 7.2.3.

13.1 Image Segmentation

Segmentation refers to a subclass of enhancement methods by which a particular object, organ, or image characteristic is extracted from the image data for purposes of visualization and measurement. Segmentation involves associating a pixel with a particular object class based on the local intensity, spatial position, neighboring pixels, or prior information about the shape characteristics of the object class. The focus of research into segmentation is to determine logic rules or strategies that accomplish acceptably accurate segmentation with as little interactive analysis as possible. Segmentation is a central problem of image analysis because it is a prerequisite for the majority of analysis methods, including image registration, shape analysis, motion detection, and volume and area estimation.

Unfortunately, there is no common method or class of methods applicable to even the majority of images. Most of the segmentation methods being applied to medical images are based on the assumption that the objects of interest have intensity or edge characteristics that allow them to be separated from the background and noise, as well as from each other. When the ranges of pixel intensities associated with different physiologic features are non-overlapping or nearly so, global thresholding (accentuating or deleting all pixels above or below a demarcating threshold of intensity) may be sufficient to provide the classification required. For example, bone is rather easily segmented from x-ray images because of the wide separation of gray-scale levels between the high-signal-intensity bone and other tissues; the intensity levels of the pixels fall into two ranges, and pixels in each can be manipulated to accentuate the difference. Global thresholding is not adequate, however, for differentiating heart muscle from chest tissues or distinguishing between cerebral gray and white matter. Classification of objects based on pixel intensity can also be implemented by use of training points, neural networks, histograms, fuzzy logic, or cluster analysis. Despite a considerable body of literature in this area, relatively few reports have demonstrated reliable segmentation based on the intensity characteristics of a single three-dimensional image.

Edge detection is simply implemented by operations that search for changes in intensity gradients; however, this is complicated because intensities in biomedical images typically ramp up or down from the structure of interest to the surrounding structure(s) from which segmentation is to be effected. Even statistical edge-finding techniques fail in most medical imaging applications where the gray-scale levels and textures of the target organ and surrounding tissues are similar—which is the usual case. Continuity and connectivity are strong criteria for separating noise from objects and have been exploited quite widely, either directly by the logic rules used in region growing¹ or by applying post-processing erosion and dilation to separate small islands of noise or spatially distinct objects with similar intensities.

A sophisticated and successful approach to segmentation is based on the spectral attributes of each pixel. LANDSAT image analysis uses the intensity corresponding to different wavelengths to differentiate regions of varying soil or vegetation content from one another. Similar approaches have been applied to MRI, wherein different pulse sequences bring out different characteristics of the magnetic resonance properties of tissues, so that cluster analysis can then be used to segment tissues with similar properties. This approach is very powerful but does require the acquisition of multiple images.

The common standard for validating or comparing segmentation methods is to examine hand-drawn contours on successive sectional scans, usually with some aid from region growing and thresholding. The goal of segmentation methods is to automate this tedious procedure, and one strategy is to begin with some model of the object. This model acts as a bound or guide to the processes, to help eliminate some ambiguity about intensity, local pixel values, or edges. The idea of using a priori knowledge about the possible shapes of objects has not been implemented in a successful segmentation system for organs, but this is a fruitful area for segmentation research.

13.2 Computational Anatomy

Organ shape analysis, including measurement of volumes and identification of surface morphology, has become feasible through modern non-invasive imaging, particularly MRI. The importance and usefulness of mathematical methods for extracting and characterizing the shape and size of organs and tumors are indicated by the richness of new information to be drawn from these measurements. Neuroanatomy is an example of a field that is beginning to employ these approaches to develop quantitative descriptors of anatomic variability across subjects, age ranges, gender, and species. Knowledge of the range of variability in normal anatomy would allow the detection and quantitative characterization of pathological deviations, for example, changes in the cerebral cortex and degeneration

¹Region growing is the process of identifying some pixels in the image that are clearly associated with different structures and then adding to each its neighboring pixels with similar intensities until regions of similar pixel intensities have been built up.

of the frontal lobe that might be linked with mental disease. If it is found that different diseases show distinctive, abnormal patterns of morphology, these patterns may ultimately provide quantitative markers for diagnosis or for assessment of the response to treatment.

Another use of organ shape analysis is to incorporate the anatomic shape and statistical properties of shape variation into algorithms for segmentation of objects that are “isointense” with respect to surrounding tissue. Indeed, although semi-automatic segmentation methods have been applied in the brain, they are often unable to deal with organs such as the heart, where image quality tends to be degraded by motion and other artifacts. The main challenge in this area is to determine which shape information is appropriate for a particular application and to make sure that the use of such prior assumptions does not bias the estimated values of model parameters. Once a shape model has been fitted to the data, dynamic properties of the object such as degree of bulging or narrowing, local deformation, and strain relationships may be derived and used to distinguish between normal and abnormal physiologic behavior.

Computational anatomy also includes the characterization of tissue architecture or surface texture, as in, for example, the analysis of changes in trabecular bone structure associated with osteoporosis. In this case, conventional methods of parameterizing the changes in image intensity have recently been extended by the use of descriptors based on Fourier space images and by fractal analyses. At present these analyses are mainly restricted to two-dimensional slices, but there is great interest in generalizing the approaches to treat three-dimensional data. Major factors influencing the results obtained with this type of analysis are the intrinsic resolution of the images and the ability to separate random noise from spatially coherent features. If these factors are properly taken into account, it is possible that textural parameters might be used to describe the microarchitecture of two- and three-dimensional image data and provide a mechanism for quantifying the severity of disease.

13.3 Registration of Multimodality Images

Registration of multimodality images is particularly important for planning surgical and radiation treatment, following changes in tissue morphology associated with disease progression or response to therapy, and relating anatomic information to changes in functional characteristics such as glucose uptake, blood flow, and cellular metabolism. The need to perform such registration is well established and has been studied quite widely for the case of registering rigid objects. The techniques that have been reported vary in detail but can be classified based on the features that are being matched. Such features include external markers that are fixed on the patient, internal anatomic markers that are identifiable on all images, the center of gravity for one or more objects in the images, crestlines of objects in the images, or gradients of intensity. One may also minimize the distance between corresponding surface points of a predefined object.

The identification of similar structures in images is a prerequisite for many image registration techniques. In some efforts this has been achieved as a manual procedure and in others by automated segmentation. When there is the possibility of tissue deformation between examinations, as is the case with soft tissue structures in the abdomen or pelvis, elastic warping is required to transform one data set into the other. The difficulty lies in defining enough common features in the images to enable specifying appropriate local deformations. Of particular interest, for example, is analysis of wall motion in the heart, which necessitates correlating positions of particular regions as a function of time in order to estimate the variations in stress and strain associated with different pathologies.

13.4 Synthesis of Parametric Images

Parametric images can be derived from any series of images using mathematical models that describe physiologic processes. Examples are flow, motion, metabolic rates, MRI relaxation times, and diffusion parameters. The motivation for producing such images is the visualization of the spatial distribution of parameters that are related to metabolic or biologically relevant tissue parameters. The ability to quantify changes in these parameters that reflect disease progression or response to therapy would be extremely valuable in assessing the effectiveness of a treatment and in providing an early indication of the need for an alternative type of therapy. Two examples of situations in which parametric images are able to provide information that is not available from conventional anatomic images are (1) in distinguishing between radiation necrosis and an active tumor through metabolite images calculated from MR spectroscopic imaging data and (2) in estimating flow parameters from velocity-encoded MR images.

The mathematical research needs for the synthesis of parametric images relate to designing parameter estimation schemes for the physiologic models being considered. The key consideration is the stability and robustness of the algorithms used to fit data at each point in space. Because of the two- or three-dimensional nature of the data, there are inevitably regions where the pixel values have very low signal-to-noise ratios. It is therefore necessary to mask out such regions or to use a fitting algorithm that behaves well for noisy data. For this reason, a wide range of different algorithms have been considered, including constrained least squares, simplex, simulated annealing, and maximum entropy.

A particularly difficult situation arises in the analysis of time series of three-dimensional MRI, positron emission tomography (PET), and single photon emission computed tomography (SPECT) data from the brain, where the differences from the baseline are significant because they represent areas of functional activation, although these differences may be on the order of just a few percent. Here, the spatial and temporal correlations may be best studied using a data reduction strategy (e.g., singular value decomposition), which could incorporate continuity and anatomic priors. A complication in the analysis of such data occurs if the subject moves significantly during the data acquisition so that successive images are no longer in correct registration.

13.5 Data Visualization

As biomedical imaging advances in terms of the sophistication of data acquisition techniques, the need to develop improved tools for image processing and visualization has become a major bottleneck. This need is particularly acute for the combined interpretation of three-dimensional anatomic and physiologic or metabolic data. The challenge is owing not only to the large size of the available data sets but also to the complexities of the relationships among the different data. One approach to this problem is data reduction or synthesis of parametric images, as has been described in previous sections. Other approaches include the use of color overlays of physiologic parameters onto anatomic structures. Such approaches are useful for making anatomic correlations but have limited scope for providing a quantitative interpretation of the relationships among different parameters.

Visualization techniques currently being investigated in computer graphics research and being applied for the analysis of biomedical data include surface rendered anatomical displays with rotation and shading, volume rendered cut-outs with enhanced emphasis of particular objects, transparent surfaces within surfaces with color shading and rotation, and reprojection techniques using various weightings of pixels of interest, such as maximum pixel intensity projection and depth weighting. An example that is already in routine use is the MRI angiogram, which is visualized in three dimensions by reprojecting at multiple different angles to form a sequence of images that can be played back in “*cine*” mode to simulate the rotation of the vessels. The design and evaluation of methods for representing biomedical image data constitute a most promising area for research, requiring a close interaction between computer scientists and the clinicians who will ultimately interpret the data.

13.6 Treatment Planning

The planning of surgical procedures, hyperthermia, cryosurgery, and radiation therapy could benefit considerably from advances in high-speed computing and image processing. The sophistication of currently used techniques derives to a large extent from improvements in the capability for acquiring volumetric images and in the interactive manipulation of those data sets within the treatment room or surgical suite. The need for immediate visualization and direct spatial correlation of structures within the body is addressed in Chapter 12 in the context of the development of interventional procedures. For hyperthermia, cryosurgery, and radiation treatment planning, an additional problem is deciding how to tailor the therapy so that the greatest possible effect is obtained within the target while the surrounding normal tissue is subjected to as low an effect as possible.

There are several different stages in planning and implementing such therapy. The first is to define the target to be treated, which in almost all cases requires the visualization of the lesion relative to the normal anatomy using diagnostic imaging modalities such as CT or MRI. Once the size and location of the lesion have been determined either by manual examination of the data or by more sophisticated image segmentation, it is necessary to determine how best to deliver the therapy. The complexity of the computations required to determine optimal delivery of therapy depends on the particular therapy used. For example, when radiation is used, it is necessary to determine the combination of radiation beam geometries, wedge placements, and dose fractionation that would be likely to deliver the maximal dose to the target with a minimal dose to surrounding tissue. This assessment requires sophisticated modeling of radiation dose distributions and may involve optimization over a very large number of different therapy plans. For tissue ablation by heating or freezing, it is necessary to determine through a model the temperature distribution that would provide the maximal effect. A demanding computation is also basic to the capability to interactively monitor the implementation of the ablation plan using a temperature-sensitive parameter such as MR relaxation time.

The computational demands associated with modern treatment planning are caused mainly by the large number of different ways in which therapy can be delivered and the time required to simulate multiple three-dimensional dose or temperature distributions. Even when the simulations can be accomplished rapidly with high-performance computing, there is still the issue of identifying a criterion to define which of the many thousands of possible options constitutes the “best” plan. Whether this identification must be achieved using interactive visual refinement to direct the optimization or whether the process can be fully automated remains to be determined for all except the simplest treatment schemes.

When the computational problem associated with defining the treatment plan has been solved, the coordinate system of the plan is then registered as accurately as possible with the patient's frame of reference. This registration is typically achieved using stereotactic frames, masks, or external markers and may need to be repeated many times in the case of temperature or dose fractionation.

13.7 Research Opportunities

There are numerous research opportunities in the field of contemporary biomedical image processing.

- Some of the most challenging research opportunities fall in the area of extending traditional approaches to segmentation and object classification in order to include shape information rather than merely image intensity. These techniques, when combined with the ability to accurately register deformable objects, would make a major contribution to interpretation of images from the heart, abdomen, and pelvis.
- A related area for research is deriving quantitative methods for analyzing tissue function and correlating that information directly to the anatomy. This research area includes the application of statistical approaches for identification of subtle changes in time series of three-dimensional images obtained for mapping brain function, the use of prior anatomic information to constrain the reconstruction of low signal-to-noise metabolic data, and the derivation of parametric images that accurately describe the kinetics of biologically relevant tracers.
- Improved techniques for visualizing multi-dimensional data are critical for establishing the relevance of new types of image data, making the information that they represent accessible to a wide audience, and understanding their relationship to conventional anatomic images.
- Radiation therapy and tissue ablation by heating or freezing require precise definition of the anatomic targets and a physical characterization of the processes leading to destruction of abnormal tissues while preserving nearby normal tissues. The need for simulation of the treatment process in three dimensions is a major challenge for high-performance computing and algorithm development.

13.8 Suggested Reading

1. Bezdek, J.C., Hall, L.O., and Clarke, L.P., Review of MR image segmentation techniques using pattern recognition, *Am. Assoc. Phys. Med.* **20** (1993), 1033–1048.
2. Bracewell, R.N., *Two-Dimensional Imaging*, Prentice Hall, Englewood Cliffs, N.J., 1995.
3. Maurer, C.R., and Fitzpatrick, J.M., A review of medical imaging registration, in *Interactive Image-Guided Neurosurgery*, R.J. Maciunas, ed., Amer. Assoc. Neurological Surgeons, Park Ridge, Ill., 17–44, 1993.
4. Robb, R.A., ed., *Visualization in Biomedical Computing 1994; Proc. SPIE*, vol. 2359, Bellingham, Wash., 1994.
5. Rosenmann, J., and Cullip, T., High performance computing in radiation cancer treatment, in *High Performance Computing in Biomedical Research*, T.C. Pilkington, B. Loftis, J.F. Thompson, S.L.-Y. Woo, T.C. Palmer, and T.F. Budinger, eds., CRC Press, Boca Raton, Fla., 465–476, 1993.
6. Thirion, J.-P., *Fast Non-Rigid Matching of 3D Medical Images*, INRIA research report 2547, INRIA, Le Chesnay, France, 1995.
7. Udupa, J.K., and Herman, G.T., eds., *3D Imaging in Medicine*, CRC Press, Boca Raton, Fla., 1991.
8. Udupa, J.K., and Samarasekera, S., Fuzzy connectedness and object definition, *Proc. Med. Imaging* **2431** (1995), 2–11.

Chapter 14

A Cross-Cutting Look at the Mathematics of Emerging Biomedical Imaging

This chapter describes the mathematical models used in medical imaging, their limitations, and the pertinent mathematical methods and problems. In many cases these problems have not yet been solved satisfactorily.

14.1 Mathematical Models for Particular Imaging Modalities

This description of mathematical models is restricted to those features that are important for the mathematical scientist. More about the physical details, the assumptions made, and the limitations can be found in Chapters 2–11.

14.1.1 Transmission Computed Tomography

Transmission computed tomography (CT) is the original and simplest case of CT. In transmission tomography one probes an object with non-diffracting radiation, e.g., x-rays for the human body. If I_0 is the intensity of the source, $a(x)$ the linear attenuation coefficient of the object at point x , L the ray along which the radiation propagates, and I the intensity past the object, then

$$I = I_0 \exp\left[-\int_L a(x) dx\right] \quad (14.1)$$

(compare to section 3.1.2). In the simplest case the ray L may be thought of as a straight line. Modeling L as a strip or cone, possibly with a weight factor to account for detector inhomogeneities, may be more appropriate. Equation 14.1 neglects the dependence of a with the energy (beam hardening effect) and other nonlinear phenomena (e.g., partial volume effect); see section 3.2.3.

The mathematical problem in transmission tomography is to determine a from measurements of I for a large set of rays L . If L is simply the straight line connecting the source x_0 with the detector x_1 , equation 14.1 gives rise to the integral

$$\log\left(\frac{I}{I_0}\right) = -\int_{x_0}^{x_1} a(x) dx, \quad (14.2)$$

where dx is the restriction to L of the Lebesgue measure in \mathbf{R}^n . The task is to compute a in a domain $\Omega \subseteq \mathbf{R}^n$ from the values of equation 14.2 where x_0 and x_1 run through certain subsets of $\partial\Omega$.

For $n = 2$, equation 14.2 is simply a reparametrization of the Radon transform R . The operator R is defined to be

$$(Ra)(\theta, s) = \int_{x \cdot \theta = s} a(x) dx, \quad (14.3)$$

where θ is a unit vector in \mathbf{R}^n ; i.e., $\theta \in S^{n-1}$, and $s \in \mathbf{R}$. Thus a is in principle found through Radon's inversion formula for R ,

$$a = R^*Kg, \quad g = Ra. \quad (14.4)$$

R^* is given explicitly by

$$(R^*g)(x) = \int_{S^{n-1}} g(\theta, x \cdot \theta) d\theta \quad (14.5)$$

and the operator K is given by

$$K = \frac{1}{2}(2\pi)^{1-n} \begin{cases} (-1)^{(n-2)/2} H \frac{\partial^{n-1}}{\partial s^{n-1}} & , \quad n \text{ even} \\ (-1)^{(n-1)/2} \frac{\partial^{n-1}}{\partial s^{n-1}} & , \quad n \text{ odd} \end{cases} \quad (14.6)$$

where H is the Hilbert transform. In fact the numerical implementation of equation 14.4 leads to the filtered backprojection algorithm, which is the standard algorithm in commercial CT scanners.

For $n = 3$, the relevant integral transform is the x-ray transform, P , given by

$$(Pa)(\theta, x) = \int_{\mathbf{R}^1} a(x + s\theta) ds ,$$

where $\theta \in S^{n-1}$ and $x \in \theta^\perp$. P admits a similar inversion formula as R :

$$a = P^*Kg , \quad g = Pf , \quad (14.7)$$

with K very similar to equation 14.6 and

$$(P^*g)(x) = \int_{S^{n-1}} g(\theta, E_\theta x) d\theta ,$$

where E_θ is the orthogonal projection onto θ^\perp . Unfortunately, equation 14.7 is not as useful as equation 14.4. The reason is that equation 14.7 requires g for all θ and $y \in \theta^\perp$; i.e., equation 14.2 has to be available for all $x_0, x_1 \in \partial\Omega$. This is not practical. Also, it is not necessary for unique reconstruction of a . In fact it can be shown that a can be recovered uniquely from equation 14.2 with sources x_0 on a circle surrounding $\text{supp}(a)$ and $x_1 \in S^{n-1}$. Unfortunately, as section 14.1.3 makes clear, the determination of a in such an arrangement is, though uniquely possible, highly unstable. The condition of stability is the following: each plane meeting $\text{supp}(a)$ must contain at least one source. This condition is obviously violated for sources on a circle. Cases in which the condition is satisfied include the helix and a pair of orthogonal circles. A variety of inversion formulas have been derived for these cases.

If scatter is to be included, a transport model is more appropriate. Let $u(x, \theta)$ be the density of the particles at x traveling (with speed 1) in direction θ . Then,

$$\theta \cdot \nabla u(x, \theta) + a(x)u(x, \theta) = \int_{S^{n-1}} \eta(x, \theta, \theta') u(x, \theta') d\theta' + \delta(x - x_0) . \quad (14.8)$$

Here, $\eta(x, \theta, \theta')$ is the probability that a particle at x traveling in direction θ is scattered in direction θ' . Again we neglect dependence on energy. δ is the Dirac δ -function modeling a source of unit strength. Equation 14.8 holds in a domain Ω of \mathbf{R}^n ($n = 2$ or 3), and $x_0 \in \partial\Omega$. Since no radiation comes in from outside we have

$$u(x, \theta) = 0 , \quad x \in \partial\Omega , \quad \nu_x \cdot \theta \leq 0 , \quad (14.9)$$

where ν_x is the exterior normal on $\partial\Omega$ at $x \in \partial\Omega$. Equation 14.1 is now replaced by

$$I(x_1, x_0, \theta) = I_0 u(x_1, \theta) , \quad x_1 \in \partial\Omega , \quad \nu_{x_1} \cdot \theta \geq 0 . \quad (14.10)$$

The problem of recovering a from equations 14.8–14.10 is much harder. An explicit formula for a such as equation 14.4 has not been found and is unlikely to exist. Nevertheless, numerical methods have been developed for special choices of η ; see the papers by Anikonov et al. and Bondarenko and Antyufeev listed in the suggested reading at the end of this chapter. The situation gets even more difficult if one takes into account that, strictly speaking, η is object dependent and hence not known in advance. Equations 14.8–14.10 are a typical example of an inverse problem for a partial differential equation. In an inverse problem one has to determine the differential equation—in this case a, η —from information about the solution—in this case equation 14.10. More on inverse problems can be found in section 14.3.

14.1.2 Emission Computed Tomography

In emission computed tomography one determines the distribution f of radiating sources in the interior of an object by measuring the radiation outside the object in a tomographic fashion. Let $u(x, \theta)$ again be the density of particles at x traveling in direction θ with speed 1, and let a be the attenuation distribution of the object. (This is the quantity that is sought in transmission CT.) Then,

$$\theta \cdot \nabla u(x, \theta) + a(x)u(x, \theta) = f(x) . \quad (14.11)$$

This equation holds in the object region $\Omega \subseteq \mathbf{R}^n$ for each $\theta \in S^{n-1}$. Again there exists no incoming radiation; i.e.,

$$u(x, \theta) = 0 , \quad x \in \partial\Omega , \quad \nu_x \cdot \theta \leq 0 , \quad (14.12)$$

while the outgoing radiation

$$u(x, \theta) = g(x, \theta) , \quad x \in \partial\Omega , \quad \nu_x \cdot \theta \geq 0 \quad (14.13)$$

is measured and hence known. Equations 14.11–14.13 again constitute an inverse problem for a transport equation. If a is known, 14.11–14.12 are readily solved to yield

$$u(x, \theta) = \int_{x-\infty \cdot \theta}^x \exp\left[-\int_y^x a \, ds\right] f(y) \, dy .$$

The exterior integral is over the ray $x - t\theta$, $0 \leq t < \infty$. Thus, equation 14.13 leads to the integral equation

$$g(x, \theta) = \int_{x-\infty \cdot \theta}^x \exp\left[-\int_y^x a \, ds\right] f(y) \, dy \quad (14.14)$$

for f ; compare to section 5.4. Apart from the exponential factor, equation 14.14 is identical—up to notation—to the integral equation in transmission CT. Except for very special cases—e.g., a constant in a known domain (see works by Tretiak and Metz and by Palamodov in the suggested reading list)—no explicit inversion formulas are available. Numerical techniques have been developed but are considered to be slow. Again the situation becomes worse if scatter is taken into account. This can be done by simply adding the scattering integral in equation 14.8 to the right-hand side of equation 14.11.

What we have described so far is called single photon emission CT (SPECT). In positron emission tomography (PET) the sources eject the particles pairwise in opposite directions. They are detected in coincidence mode; i.e., only events with two particles arriving at opposite detectors at the same time are counted. Equation 14.14 has to be replaced (see section 6.1.3) by

$$\begin{aligned} g(x, \theta) &= \int_{x-\infty \cdot \theta}^x \exp\left[-\int_y^x a \, ds - \int_{x-\infty \cdot \theta}^y a \, ds\right] f(y) \, dy \\ &= \exp\left[-\int_{x-\infty \cdot \theta}^x a \, ds\right] \int_{x-\infty \cdot \theta}^x f(y) \, dy . \end{aligned} \quad (14.15)$$

Thus PET is even closer to the case of transmission CT. If a is known, we simply have to invert the x-ray transform. Inversion formulas are available that can make use of the data collected in PET, and their numerical implementations are currently under consideration.

The main problems in emission CT are (compare to sections 5.4.2 and 6.2.4) unknown attenuation, noise, and scatter. For the attenuation problem, the ideal mathematical solution would be a method for determining f and a in equations 14.11–14.13 simultaneously. Under strong assumptions on a — a constant in a known region (Hertle's paper in section 14.8), a an affine distortion of a prototype (Natterer's 1993 paper on tissue attenuation in section 14.8), or a close to a known distribution (Bronnikov's paper in section 14.8)—encouraging results have been obtained. Theoretical results based on the transport formulation have been derived, even for models including scatter (see Romanov's work listed in section 14.8). But a clinically useful way of determining a from the emission data has not yet been found.

Noise and scatter are stochastic phenomena. Thus, in addition to models using integral equations, stochastic models have been set up for emission tomography. These models are completely discrete. The reconstruction region is subdivided into m pixels or voxels. The number of events in pixel/voxel j is a Poisson random variable φ_j whose mathematical expectation $f_j = E\varphi_j$ is a measure for the activity in pixel/voxel j . The vector $f = (f_1, \dots, f_n)$ is the sought-for quantity. The vector $g = (g_1, \dots, g_n)$ of measurements is considered a realization of the random variable $\gamma = (\gamma_1, \dots, \gamma_n)$, where γ_i is the number of events detected in detector i . The model is determined by the (n, m) -matrix $A = (a_{ij})$ whose elements are

$$a_{ij} = P(\text{event in pixel/voxel } j \text{ is detected in detector } i) ,$$

where P denotes probability. Normalizing the a_{ij} such that $\sum_{j=1}^m a_{ij} = 1$ leads to $E(\gamma) = Af$. f is determined from g by the maximum likelihood method. A numerical method for doing this is the expectation maximization (EM) algorithm. In its basic form it reads (compare section 6.2.4)

$$f^{k+1} = f^k A^* \frac{g}{Af^k} , \quad k = 0, 1, \dots ,$$

where division and multiplication are to be understood componentwise. The problem with the EM-algorithm is that it is only semi-convergent (compare section 14.4); i.e., noise is amplified at high iteration numbers. This is known as the checkerboard effect. Various suggestions have been made for getting rid of this effect. The most exciting and interesting ones use prior information and attempt to maximize “posterior likelihood.” Thus f is assumed to have a prior probability distribution, called a Gibbs-Markov random field $\pi(f)$, which gives preference to certain functions f . Most priors π simply add a penalty term to the likelihood function to account for correction between neighboring pixels and do not use biological information. However, if π is carefully chosen so that piecewise constant functions f with smooth boundaries forming the region of constancy are preferred, then the noise amplification at high iteration numbers can be avoided. The question remains as to whether this conclusion will remain valid for functions f that are assigned low probability by π —or, more to the point, whether “real” emission densities f will be well-resolved by this Bayesian method. A double-blind study in which radiologists tried to find lesions from images produced by two different algorithms concluded that maximum likelihood methods were superior to the filtered backprojection algorithm in certain clinical applications. The same type of study is needed to determine whether or not Gibbs priors will improve the maximum likelihood reconstruction (stopped short of convergence to avoid noise amplification) on real data.

14.1.3 Ultrasound Computed Tomography

X-rays travel along straight lines. For other sources of radiation, such as ultrasound and microwaves, this is not the case. The paths are not straight, and their exact shape depends on the internal structure of the object. Simple projections and linear integral equations will not suffice, and more sophisticated nonlinear models have to be used.

In the following we consider an object $\Omega \subseteq \mathbf{R}^n$ with refractive index γ . Assume $\gamma = 1$ outside the object. The object is probed by a plane wave

$$e^{-ikt} u_\theta(x) , \quad u_\theta(x) = e^{ikx \cdot \theta}$$

with wave number $k = \frac{2\pi}{\lambda}$, λ the wave length, traveling in the direction θ . The resulting wave $e^{-ikt} u(x)$ satisfies the reduced wave equation

$$\Delta u + k^2(1 + f)u = 0 , \quad f = \gamma^2 - 1 , \quad (14.16)$$

plus suitable boundary conditions at infinity. The inverse problem to be solved is now the following. Assume that

$$g(x, \theta) = u(x) , \quad \theta \in S^{n-1} \quad (14.17)$$

is known outside Ω . Determine f inside Ω !

The uniqueness and stability of the inverse problem equations 14.16–14.17 have recently been settled. However, stability is only logarithmic; i.e., a data error of size δ results in a reconstruction error of $1/\log(1/\delta)$ (see section 14.4). Numerical algorithms did not emerge from this work.

Numerical methods for equations 14.16–14.17 are based mostly on linearizations, such as the Born and Rytov approximations. In order to derive the Born approximation, one rewrites 14.16 as

$$u(x) = u_\theta(x) - k^2 \int_{\Omega} G(x-y) f(y) u(y) dy , \quad (14.18)$$

where G is an appropriate Green’s function. For $n = 3$, the Green’s function is

$$G(x) = \frac{e^{ik|x|}}{4\pi|x|} . \quad (14.19)$$

The Born approximation is now obtained by assuming $u \sim u_\theta$ in the integral in equation 14.18. With this approximation, equation 14.17 reads

$$g(x, \theta) = u_\theta(x) - k^2 \int_{\Omega} G(x-y) u_\theta(y) f(y) dy , \quad x \notin \Omega . \quad (14.20)$$

This is a linear integral equation for f , valid for all x outside the object and for all measured directions θ .

Numerical methods based on equation 14.20—and a similar equation for the Rytov approximation—have become known as diffraction tomography. Unfortunately, the assumptions underlying the Born and Rytov approximations are not satisfied in medical imaging. Thus, the reconstructions of f obtained from equation 14.20 are very poor. However, we may use equation 14.20 to get some encouraging information about stability. For $|x|$ large, equation 14.20 assumes the form

$$\begin{aligned} g(x, \theta) &= u_\theta(x) - \frac{k^2}{4\pi|x|} e^{ik|x|} \int e^{ik(\theta - \frac{x}{|x|}) \cdot y} f(y) dy \\ &= u_\theta(x) - \frac{k^2}{4\pi|x|} e^{ik|x|} (2\pi)^{3/2} \hat{f}[k(\frac{x}{|x|} - \theta)], \end{aligned} \quad (14.21)$$

with \hat{f} the Fourier transform of f . Equation 14.21 determines \hat{f} within a ball of radius $\sqrt{2}k$ from the data in equation 14.17 in a completely stable way. Therefore, the stability of the inverse problem 14.16–14.17 is much better than logarithmic. If the resolution is restricted to spatial frequencies below $\sqrt{2}k$ —which is perfectly reasonable from a physical point of view—then we can expect equations 14.16–14.17 to be perfectly stable.

So far we have considered plane wave irradiation at fixed frequency, and have worked in the frequency domain. Time domain methods are conceivable as well. We start with the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \Delta u, \quad (14.22)$$

with the propagation speed c assumed to be 1 outside the object. With x_0 a source outside the object, consider the initial conditions

$$u(x, 0) = 0, \quad \frac{\partial u}{\partial t}(x, 0) = \delta(x - x_0). \quad (14.23)$$

We want to determine c inside the object from knowing

$$g(x_0, x_1, t) = u(x_1, t), \quad t > 0 \quad (14.24)$$

for many sources x_0 and receivers x_1 outside the object. In the one-dimensional case, the inverse problem described by equations 14.22–14.24 can be solved by the famous Gelfand-Levitan method in a stable way. It is not clear how Gelfand-Levitan can be extended to dimensions two and three. The standard methods use sources and receivers on all of the boundary of the object. This is not practical in medical imaging. However, for reduced data sets, comparable to those in three-dimensional x-ray tomography (compare equation 14.1), it is not known how to use Gelfand-Levitan, nor is anything known about stability.

Of course one can always solve the nonlinear problem of equations 14.22–14.24 by a Newton-type method. Such methods have been developed (see papers by Gutman and Klivanov and by Kleinman and van den Berg in section 14.8). They suffer from the requirement for excessive computing time and from their apparent inability to handle large wave numbers k .

14.1.4 Optical Tomography

With optical tomography one uses near-infrared lasers for the illumination of the body (see Chapter 11). The process is described by the transport equation

$$\begin{aligned} \frac{\partial u}{\partial t}(x, \theta, t) + \theta \cdot \nabla u(x, \theta, t) + a(x)u(x, \theta, t) \\ = b(x) \int_{S^{n-1}} \eta(\theta \cdot \theta') u(x, \theta', t) d\theta' + f(x, \theta, t) \end{aligned} \quad (14.25)$$

for the density $u(x, \theta, t)$ of the photons at $x \in \Omega$ flying in direction $\theta \in S^{n-1}$ at time t . a and b are the sought-for tissue parameters. In the simplest case the scattering kernel η is assumed to be known. The source term f is under the control of the experimenter. Together with the initial and boundary conditions

$$\begin{aligned} u(x, \theta, 0) &= 0 \quad \text{in } \Omega \times S^{n-1}, \\ u(x, \theta, t) &= 0 \quad \text{on } \partial\Omega \times S^{n-1} \times \mathbf{R}^1, \quad \nu_x \cdot \theta \leq 0, \end{aligned} \quad (14.26)$$

equation 14.25 has a unique solution under natural conditions on a , b , η , and f . As in equation 14.1, we pose the inverse problem: assume that the outward radiation is known,

$$g(x, \theta, t) = u(x, \theta, t) \quad \text{on } \partial\Omega \times S^{n-1} \times \mathbf{R}^1, \quad \nu_x \cdot \theta \geq 0; \quad (14.27)$$

can we determine one or both of the quantities a , b ?

There are essentially three methods for illuminating the object, i.e., for choosing the source term f in equation 14.25. In the stationary case one posits $f = \delta(x - x_0)$, where $x_0 \in \partial\Omega$ is a source point. u is considered stationary, too. A second possibility is the light flash $f = \delta(x - x_0)\delta(t)$. Finally, one can also use time harmonic illumination, in which case $f = \delta(x - x_0)e^{i\omega t}$. This case reduces to the stationary case with a replaced by $a + i\omega$. In all three cases, the data function g of equation 14.27 is measured at $x \in \partial\Omega$, possibly averaged over one or both of the variables θ , t .

Light tomography is essentially an absorption and scattering phenomenon. This means that the scattering integral in equation 14.25 is essential; it can no longer be treated merely as a perturbation, as in x-ray CT. Thus the mathematical analysis and the numerical methods are expected to be quite different from what is seen in other types of tomography.

The mathematical theory of the inverse problem 14.25–14.27 is in a deplorable state. There exist some Russian papers on uniqueness (see Anikonov in section 14.8). General methods have been developed, too, but apparently they have been applied to one-dimensional problems only (see Sanchez and McCormick in section 14.8). Nothing seems to be known about stability. The numerical methods that have become known are of the Newton type, either applied directly to the transport equation or to the so-called diffusion approximation (e.g., Arridge et al., listed in section 14.8). The diffusion approximation is an approximation to the transport equation by a parabolic differential equation. Since inverse problems for parabolic equations are severely ill posed, this approach is questionable. Higher-order approximations (see, e.g., Gratton et al., in section 14.8) are hyperbolic, making the inverse problem much more stable.

As an alternative to the transport equation one can also model optical tomography by a discrete stochastic model. We consider only a very simple model of the two-dimensional case. Break up the object into a rectangular arrangement of pixels labeled by indices i, j with $a \leq i \leq b$ and $c \leq j \leq d$. Attach to each pixel the quantities f_{ij} , b_{ij} , r_{ij} , ℓ_{ij} denoting the probability (respectively) of a forward, backward, rightward, or leftward transition out of the pixel $\{i, j\}$ with respect to the direction used to get into this pixel. For each pair of boundary pixels $\{i, j\}$ and $\{i', j'\}$, let $P_{ij, i'j'}$ be the probability that a particle that enters the object at pixel $\{i, j\}$ will eventually leave the object at pixel $\{i', j'\}$. The problem is to determine the quantities f_{ij} , b_{ij} , r_{ij} , ℓ_{ij} from the values of $P_{ij, i'j'}$ for all boundary pixels. Preliminary numerical tests show that this is possible, at least in principle. However, the computations are very time-consuming. More seriously, they reveal a very high degree of instability.

14.1.5 Electrical Impedance Tomography

Here, the sought-for quantity is the electrical impedance σ of an object Ω . Voltages are applied via electrodes on $\partial\Omega$, and the resulting currents at these electrodes are measured. With u the potential in Ω , we have

$$\begin{aligned} \operatorname{div}(\sigma \cdot \nabla u) &= 0 \quad \text{in } \Omega \\ u &= g, \quad \sigma \frac{\partial u}{\partial \nu} = f \quad \text{on } \partial\Omega. \end{aligned} \quad (14.28)$$

Knowing many voltage-current pairs (g, f) on $\partial\Omega$, the task is to determine σ from equation 14.28.

Uniqueness for the inverse problem 14.28 has recently been settled. Unfortunately, the stability properties are very bad. Numerical methods based on Newton's method, linearization, simple backprojection, and layer stripping have been tried. All these methods suffer from the severe ill-posedness of the problem. There seems to be no way to improve stability by purely mathematical means.

14.1.6 Magnetic Resonance Imaging

The physical phenomenon exploited in magnetic resonance imaging (MRI) is the precession of the spin of a proton in a magnetic field of strength H about the direction of that field. The frequency of this precession is the Larmor frequency γH where γ is the gyromagnetic ratio. By making the magnetic field H space-dependent in a controlled way, the local magnetization $M_0(x)$ (together with the relaxation times $T_1(x)$ and $T_2(x)$) can be imaged. In the following we derive the imaging equations; compare to section 4.1.

The magnetization $M(x, t)$ caused by a magnetic field $H(x, t)$ satisfies the Bloch equation

$$\frac{\partial M}{\partial t} = \gamma M \times H - \frac{1}{T_2}(M_1 e_1 + M_2 e_2) - \frac{1}{T_1}(M_3 - M_0) e_3. \quad (14.29)$$

Here, M_i and M_i^0 are the i -th components of M and M^0 , respectively, and e_i is the i -th unit vector $i = 1, 2, 3$. The significance of T_1 , T_2 , and M_0 becomes apparent if equation 14.29 is solved with the static field $H = H_0 e_3$ and initial values $M(x, 0) = M^0(x)$. Setting $\omega_0 = \gamma H_0$ leads to

$$\begin{aligned} M_1(x, t) &= e^{-t/T_2}(M_1^0 \cos(\omega_0 t) + M_2^0 \sin(\omega_0 t)) \\ M_2(x, t) &= e^{-t/T_2}(-M_1^0 \sin(\omega_0 t) + M_2^0 \cos(\omega_0 t)) \\ M_3(x, t) &= e^{-t/T_1} M_3^0 + (1 - e^{-t/T_1}) M_0. \end{aligned} \quad (14.30)$$

Thus the magnetization rotates in the (x_1, x_2) -plane with Larmor frequency ω_0 and returns to the equilibrium condition $(0, 0, M_0)$ with speed controlled by T_2 in the (x_1, x_2) -plane and by T_1 in the x_3 -direction.

In an MRI scanner one generates a field

$$H(x, t) = (H_0 + G(t) \cdot x) e_3 + H_1(t)(\cos(\omega_0 t) e_1 + \sin(\omega_0 t) e_2),$$

where G and H_1 are under control. In the language of MRI, $H_0 e_3$ is the static field, G the gradient, and H_1 the radio-frequency (RF) field. The input G , H_1 produces in the detecting system of the scanner the output signal

$$S(t) = -\frac{d}{dt} \int_{\mathbf{R}^3} M(x, t) B(x) dx, \quad (14.31)$$

where B characterizes the detecting system. Depending on the choice of H_1 , various approximations to S can be derived, two of which are detailed here.

Case 1, short $\pi/2$ pulse In the first case, H_1 is constant in the small interval $[0, \tau]$ and $\gamma \int_0^\tau H_1 dt = \frac{\pi}{2}$ (short $\frac{\pi}{2}$ pulse). In that case,

$$S(t) = \int_{\mathbf{R}^3} M_0(x) e^{-i\gamma \int_0^t G(t') dt' \cdot x - t/T_2(x)} dx.$$

Choosing G constant for $\tau \leq t \leq \tau + T$ and zero otherwise, we get for $T \ll T_2$

$$\begin{aligned} S(t) &= \int_{\mathbf{R}^3} M_0(x) e^{-i\gamma(t-\tau)G \cdot x} dx \\ &= (2\pi)^{3/2} \hat{M}_0(\gamma(t-\tau)G), \end{aligned} \quad (14.32)$$

where \hat{M}_0 is the three-dimensional Fourier transform of M_0 .

From here we can proceed in two ways. We can use equation 14.32 to determine the three-dimensional Fourier transform \hat{M}_0 of M_0 and compute M_0 by an inverse three-dimensional Fourier transform. This requires \hat{M}_0 to be known on a cartesian grid, which can be achieved by a proper choice of the gradients or by interpolation. Alternatively, one can invoke the central slice theorem to obtain the three-dimensional Radon transform RM_0 of M_0 by a series of one-dimensional Fourier transforms. M_0 is then recovered by inverting the three-dimensional Radon transform.

Case 2, shaped pulse In this case, H_1 is the shaped pulse

$$H_1(t) = \phi(t\gamma G) e^{i\gamma G x_3 t},$$

where ϕ is a smooth positive function supported in $[0, \tau]$. Then, with x' , G' the first two components of x , G , respectively, we have

$$S(t) = \int_{\mathbf{R}^2} M'_0(x', x_3) e^{-i\gamma \int_0^t G'(t') dt' \cdot x' - t/T_2(x', x_3)} dx', \quad (14.33)$$

where

$$M'_0(x', x_3) = \int M_0(x', y_3) Q(x'_3 - y_3) dy_3,$$

with a function Q essentially supported in a small neighborhood of the origin. Equation 14.33 is the two-dimensional analog of equation 14.32, and again we face the choice between Fourier imaging (i.e., computing the two-dimensional Fourier transform from 14.33 and doing an inverse two-dimensional Fourier transform) and projection imaging (i.e., doing a series of one-dimensional Fourier transforms on equation 14.33 and inverting the two-dimensional Radon transform).

Some of the mathematical problems of MRI, e.g., interpolation in Fourier space, are common to other techniques in medical imaging and are discussed in section 14.5. An interesting mathematical problem occurs if the magnets do not produce sufficiently homogeneous fields. This then calls for the reconstruction of a function from integrals over slightly curved manifolds. Even though this is a problem of classical integral geometry, a satisfactory solution has not yet been found.

14.1.7 Vector Tomography

If the domain under consideration contains a moving fluid, then the Doppler shift can be used to measure the velocity $u(x)$ of motion. Assume the time-harmonic signal $e^{i\omega_0 t}$ is transmitted along the oriented line L . This signal is reflected by particles traveling with speed ν in the direction of L according to $e^{i(\omega_0 - k\nu)t}$, where $k = 2\omega_0/c$ and c is the speed of the probing signal; thus, $k\nu$ is the Doppler shift. Let $S(L, \nu)$ be the Lebesgue measure of those particles on L that move with speed less than ν ; i.e., $u(x) \cdot e_L < \nu$, where e_L is the tangent vector on L . Then the total response is

$$g(L, t) = \int_{-\infty}^{+\infty} e^{i(\omega_0 - k\nu)t} S(L, \nu) d\nu .$$

Thus S can be recovered from g by a Fourier transform. The problem is to recover u from S .

Not much is known about uniqueness of the solution. However, the first moment of S ,

$$\int_{-\infty}^{+\infty} \nu S(L, \nu) d\nu = \int_L u(x) \cdot e_L dx = (Ru)(L), \tag{14.34}$$

is similar to the Radon transform from section 14.1.1. Inverting R is a problem of vector tomography. Mathematically, this problem belongs to the recently developed field of integral geometry of tensor fields; see, for example, the paper by Sharafutdinov listed in section 14.8. One can show that curl u can be computed from Ru , and an inversion formula similar to the Radon inversion formula exists. Numerical simulations of Doppler tomography are given in the work by Sparr et al. listed in section 14.8.

14.1.8 Tensor Tomography

As an immediate extension of transmission CT to anisotropic media, consider a matrix-valued attenuation $a(x) = (a_{ij}(x))$, $i, j = 1, \dots, n$. We then solve the vector differential equation

$$\frac{du(t)}{dt} = -a[x(t)]u(t), \quad x(t) = (1-t)x_0 + tx_1, \quad 0 \leq t \leq 1 \tag{14.35}$$

for the vector-valued function $u(t) = [u_i(t)]_{i=1, \dots, n}$. Let a be defined in a convex domain Ω , and let $x_0, x_1 \in \partial\Omega$. Then, $u(x_1) = U(x_0, x_1)u(x_0)$ with a nonlinear map $U(x_0, x_1)$ depending on a . The problem is to recover a in Ω from the knowledge of $U(x_0, x_1)$ for $x_0, x_1 \in \partial\Omega$. For $n = 1$ we regain equation 14.1. Applications of equation 14.35 for $n > 1$ have become known in photoelasticity (see Aben's book, listed in the suggested reading), but applications to medicine are not totally out of the question.

In a further extension, let a depend on the direction $\xi = (x_1 - x_0)/|x_1 - x_0|$. Such situations occur, for instance, in the polarization of harmonic electromagnetic and elastic waves in anisotropic media. In linearized form these problems give rise to the transverse x-ray transform

$$(Ja)(x, \theta, \omega) = \int \omega^T a(x + t\theta)\omega dt, \quad \omega \perp \theta \tag{14.36}$$

and to the longitudinal x-ray transform

$$(Ja)(x, \theta) = \int \theta^T a(x + t\theta)\theta dt. \tag{14.37}$$

Equation 14.36 can easily be reduced to the $(n - 1)$ -dimensional x-ray transform in the plane $H_{\omega, s} = \{y : y \cdot \omega = s\}$. We only have to introduce the function $a_{\omega}(y) = \omega^T a(y)\omega$. Then, 14.36 provides for $x \in H_{\omega, s}$ all the line integrals of

a_ω on $H_{\omega,s}$. For equation 14.37, the situation is not so easy. We decompose a into its solenoidal and potential parts; i.e.,

$$a = a_1 + \nabla a_2, \quad \operatorname{div} a_1 = 0 \quad a_2 = 0 \quad \text{on } \partial\Omega .$$

It can be shown that a_1 can be recovered uniquely from 14.37, but a_2 is completely undetermined. This is reminiscent of vector tomography as discussed in subsection 14.1.7.

14.1.9 Magnetic Source Imaging

In magnetic source imaging (MSI) one wants to find the electric currents inside the body from measurements of the induced magnetic field outside the patient. These fields may be caused by epileptic fits or by cardiac infarction. The appropriate mathematical model relating the electric field J inside the body Ω to the magnetic field B is the Biot-Savart law (compare section 10.2)

$$B(x) = \frac{\mu_0}{4\pi} \int_{\Omega} J(y) \times \frac{x-y}{|x-y|^3} dy . \tag{14.38}$$

Here, $J = J^i + \sigma E$ with J^i the impressed source current and σE the ohmic current. Since σE depends on J^i , 14.38 is a nonlinear equation for J^i .

In general the relation between J^i and σE is quite complicated and requires extensive computations (see the work by Hämäläinen and Sarvas listed in section 14.8). If the body is modeled as a spherically symmetric conductor, this relation can be made explicit. In that case, 14.38 reduces to the linear integral equation

$$B(x) = \frac{\mu_0}{4\pi} \int_{\Omega} \left\{ \frac{y \times Q(y)}{F(x,y)} - \frac{(y \times Q(y)) \cdot x}{F^2(x,y)} \nabla_x F(x,y) \right\} dy , \tag{14.39}$$

$$F(x,y) = |x-y|(|x-y||x| + (x-y) \cdot x) ,$$

for the dipole moment Q . Obviously, equation 14.39 is not uniquely solvable for Q . However, it makes sense to compute the generalized solution defined by 14.47 below, which is called the lead field solution in the biomagnetism literature.

Besides non-uniqueness, the main problem in MSI is again instability. One tries to overcome this difficulty either by assuming that there are only a few (1-3) magnetic dipoles or by invoking a strong regularization scheme for a continuous dipole distribution.

14.1.10 Electrical Source Imaging

With electrical source imaging (ESI) one wants to find the electrical potential u generated by electrical sources inside the body (typically in the heart) on a surface Γ_1 (“epicardial”) close to the sources from the potential on the surface Γ_0 of the body Ω . Thus we have (see also section 8.2)

$$\begin{aligned} \Delta u &= 0 & \text{between } \Gamma_1 & \text{ and } \Gamma_0 \\ \frac{\partial u}{\partial \nu} &= 0 & \text{on } \Gamma_0 . \end{aligned} \tag{14.40}$$

The problem is to recover u on Γ_1 from the values of u on a part of Γ_0 . The degree of ill-posedness of this inverse problem depends on how close Γ_1 is to the sources. In an idealized setting this can be analyzed quantitatively as was done by Lavrentiev et al. (see suggested reading list). Let Γ_1, Γ_0 be concentric circles with radius $r_1 < r_0$, respectively, and let all the sources lie inside a third concentric circle with radius $r_2 < r_1$. Then, the reconstruction error $\varepsilon(\delta)$ on Γ_1 for a data error δ on Γ_0 is

$$\varepsilon(\delta) = M_2^{1-c} \delta^c, \quad \text{where} \tag{14.41}$$

$$c = \frac{\log(r_1/r_2)}{\log(r_0/r_2)}, \quad M_2 = \max_{|x|=r_2} |u(x)| .$$

In the terminology of section 14.4, the problem is only modestly ill posed if c is not too small, i.e., if the sources are not too close to Γ_1 .

Unfortunately, in practice Γ_1 is always very close to the sources; i.e., c is small. Thus the inverse problem 14.40 behaves like a severely ill posed problem, and regularization is mandatory. Let $A : L_2(\Gamma_1) \rightarrow L_2(\Gamma_0)$ be the forward

operator, which associates with each potential f on Γ_1 the function $g = u$ on Γ_0 where u is the solution of equation 14.40 with $u = f$ on Γ_1 . Then, the Tikhonov-Phillips regularization is obtained by minimizing

$$\|Af - g\|_{L_2(\Gamma_0)}^2 + \alpha^2 \|f - f_0\|_{L_2(\Gamma_1)}^2.$$

A possible choice for f_0 is $f_0 = 0$. However, since measurements are done for many time instants, one can also use an approximation to f computed from earlier time frames.

14.2 Forward Problems

Many of the equations described above are derived as approximations to complex physical phenomena. A good formulation in terms of the tissue parameters relative to each modality is the basis of any attack on the corresponding inverse problem. Considerable effort should be spent producing increasingly realistic models that allow for effective numerical simulation and validation in terms of real data.

Numerical methods for solving inverse problems are usually iterative in nature—solving the forward problem over and over again. Thus it is important to have fast and reliable algorithms for the forward problems. Such algorithms are available for impedance tomography, since algorithms for Poisson-type equations are a well-established field in numerical analysis (multigrid methods, domain decomposition, preconditioning). For ultrasound, one has to solve Helmholtz-type equations at high spatial frequencies, which still is a challenge. Likewise, the development of codes for transport problems is still an area of active research.

14.3 Inverse Problems

Most of the mathematical problems in medical imaging belong to the class of inverse problems. In an abstract framework these problems can be described as follows. Let $L(a)$ be an operator depending on the parameter a . Assume the equation

$$L(a)u = f \tag{14.42}$$

to be solvable for each f . Equation 14.42 is called the forward problem in this context. Now assume we are given an “observation operator” B , and let

$$Bu = g \tag{14.43}$$

be known. The inverse problem calls for computing a from the overdetermined system 14.42–14.43. A typical example is the ultrasound problem discussed in section 14.1.3, in which $L(a)$ is the Helmholtz operator $\Delta + k^2(1+a)$ or the hyperbolic operator $\partial^2 u / \partial t^2 - c^2 \Delta u$. In transmission and optical tomography, $L(a)$ is a transport operator. For impedance tomography the operator is elliptic, $L(a) = \text{div}(a \nabla u)$. Parabolic operators show up in the diffusion approximation to optical tomography.

Very often the inverse problem 14.42–14.43 can be reduced either exactly or approximately to a linear integral equation of the first kind,

$$\int_{\Omega} K(x, y) a(y) dy = g(x), \quad x \in M. \tag{14.44}$$

Depending on the smoothness of the kernel K , 14.44 is more or less ill posed. We come back to this point in section 14.4.

Since the inverse operator $L^{-1}(a)$ exists, the inverse problem 14.42–14.43 is equivalent to the single nonlinear equation

$$BL^{-1}(a) = g. \tag{14.45}$$

In principle, every method for solving nonlinear ill-posed problems (such as a properly regularized Newton method or the Levenberg-Marquardt method of section 14.4) can be used for 14.42–14.43. However, it is clear that the structure of the problem at hand has to be exploited. For instance, in all the cases discussed here, $L(a)$ depends in an affine-linear way on a . Also, $L(a)$ decomposes in many cases naturally into lower-dimensional operators $L_j(a)$, $j = 1, \dots, p$, of identical structure. For instance, j may refer to the direction of incidence in x-ray/ultrasound tomography or to the position of the electrode in impedance tomography.

The mathematical theory of inverse problems is still in its infancy. Uniqueness theorems have been obtained only recently. However, the most serious problem is instability. Integral equations of the first kind with smooth kernels, such as 14.44, tend to be severely ill posed; see the next section. The few exceptions to this rule include inverse problems for hyperbolic equations. In particular in the one-dimensional case, the Gelfand-Levitán procedure not only yields uniqueness and stability, but also provides efficient numerical methods (background can be found in the work by Burridge listed in section 14.8).

14.4 Ill-Posedness and Regularization

All problems in medical imaging are ill posed, even to very different degrees. We describe ill-posedness for linear equations such as 14.44 and discuss some widely used regularization techniques. The results are extended to nonlinear ill-posed problems. Discussion of the use of nonlinear and stochastic side information is postponed to section 14.6.

Let A be a map from a Hilbert space X into a Hilbert space Y . The equation

$$Af = g \tag{14.46}$$

is called well posed if

- (a) the equation is solvable for each g ,
- (b) the solution f is uniquely determined, and
- (c) the solution f depends continuously on g .

Otherwise, equation 14.46 is ill posed. The (Moore-Penrose) generalized solution f^+ of 14.46 is defined in the following way: among all the least squares solutions, i.e., the minimizers of $\|Af - g\|$, select the one for which $\|f\|$ is minimal. If A is linear and bounded, f^+ can be characterized by

$$A^*Af^+ = A^*g, \quad f^+ = A^*u, \quad u \in Y. \tag{14.47}$$

f^+ is uniquely determined. In many cases in medical imaging, f^+ is a very useful substitute for the exact solution of equation 14.46.

In medical imaging, the failure to satisfy condition (c) is quite common, which leads to instability. In that case, a small data error δ of g can lead to a very large error in the computed solution f .

Stability can be restored by side information. By this we mean information on f that is not contained in 14.46, such as $f \in M$ where M is a subset of X . Often one can find a function ε with $\varepsilon(\delta) \rightarrow 0$ as $\delta \rightarrow 0$ such that the following (conditional) stability estimate holds: let $g^\delta \in Y$ such that $\|g - g^\delta\| \leq \delta$, and let $Af^\delta = g^\delta$. Let $f \in M$. Then,

$$\|f - f^\delta\| \leq \varepsilon(\delta). \tag{14.48}$$

The behavior of $\varepsilon(\delta)$ as $\delta \rightarrow 0$ reflects the degree of ill-posedness. If $\varepsilon(\delta)$ is of the order δ , then there is no ill-posedness at all. On the other hand, if $\varepsilon(\delta)$ tends to zero very slowly, e.g., $\varepsilon(\delta) = 1/\log(1/\delta)$ (logarithmic stability), then the problem is severely ill posed. For such problems, reducing the reconstruction error by a factor of 2 requires doubling the number of correct digits in the data. Such problems are extremely difficult to handle. As a rule, only very poor approximations to the true solution can be computed. The only problems in medical imaging that have been solved successfully so far belong to the class of modestly ill-posed problems, for which $\varepsilon(\delta)$ is of the order δ^c with $0 < c < 1$ (Hölder stability). For this class, m digit accuracy in the data leads to cm correct digits in the reconstruction. This loss in accuracy seems bearable if c is not too small.

A regularization method is any method that computes from g^δ and from a knowledge of M an element f^δ such that 14.48 holds. Some common methods are explored below.

14.4.1 The Tikhonov-Phillips Method

With the Tikhonov-Phillips method of regularization (see also section 8.2.2), $M = \{f \in X : \|f - f_0\| \leq \rho\}$ for some $f_0 \in X$, $\rho > 0$. f^δ is obtained by minimizing

$$\|Af - g^\delta\|^2 + \alpha^2\|f - f_0\|^2 \tag{14.49}$$

in X . If A is linear, then f^δ can be computed by solving the linear system

$$(A^*A + \alpha^2I)(f^\delta - f_0) = A^*(g^\delta - Af_0) \tag{14.50}$$

or, equivalently,

$$f^\delta = f_0 + A^*(AA^* + \alpha^2I)^{-1}(g^\delta - Af_0), \tag{14.51}$$

where A^* is the adjoint of A . The choice of the regularization parameter α is critical. Of course it depends primarily on $\delta : \alpha = \alpha(\delta)$.

14.4.2 The Truncated Singular Value Decomposition

A is said to admit a singular value decomposition (SVD) if

$$\begin{aligned} Af &= \sum_{k=1}^{\infty} \sigma_k(f, f_k) g_k, \\ A^* f &= \sum_{k=1}^n \sigma_k(g, g_k) f_k, \end{aligned}$$

where (f_k) , (g_k) are orthonormal systems in X , Y , respectively, and σ_k are positive numbers, the singular values of A . The spectrum of A^*A consists of the numbers σ_k^2 and possibly the eigenvalue 0. The decay of the singular values is a measure for the degree of ill-posedness. Exponential decay corresponds to severe ill-posedness.

The SVD provides more information on the character of the ill-posedness than does any other method. It tells us precisely which features of f cannot be recovered in a stable way, namely those that resemble functions f_k that have small singular values σ_k . A good example is the limited angle problem in transmission CT where the SVD has been computed analytically.

With the help of SVD one can compute an approximation f^δ to f by

$$f^\delta = \sum_{k, \sigma_k \geq \sigma} \frac{1}{\sigma_k} (g^\delta, g_k) f_k, \quad (14.52)$$

where $\sigma = \sigma(\delta)$ plays the role of a regularization parameter. In most cases in medical imaging, the SVD is too expensive to provide an efficient practical reconstruction method. But for analyzing a problem it is a very valuable tool.

14.4.3 Iterative Methods

Let $f_{k+1} = B_k f_k + C_k g$ be any convergent iterative method for the solution of $Af = g$. When applied to g^δ instead of g , the sequence will in general not converge: even if the first few iterates provide a good approximation to f , they will deteriorate as the iteration goes on. This semi-convergence phenomenon has been observed in many iterative methods in medical imaging, most prominently in the EM algorithm in emission tomography (checkerboard effect, see page 95).

Semi-convergence forces one to stop the iteration after a certain number $k(\delta)$ of steps. We put

$$f^\delta = f_{k(\delta)}. \quad (14.53)$$

The optimal choice of $k(\delta)$ is the subject of many papers.

14.4.4 Regularization by Discretization

Let A_h be a discrete approximation to A with step size h . For $h \rightarrow 0$ we assume $A_h \rightarrow A$ in some sense. Since A^{-1} is not continuous, A_h^{-1} (if it exists) tends to infinity in some sense. Thus,

$$f_h = A_h^{-1} g_h,$$

with some discrete approximation g_h to g , will not converge, certainly not when the procedure is applied to g^δ instead of g . However, if we choose $h = h(\delta)$ suitably, then, subject to certain assumptions,

$$f^\delta = A_{h(\delta)} g_{h(\delta)} \quad (14.54)$$

will satisfy equation 14.48. Often the discretization is done by projecting on finite dimensional subspaces. In the medical imaging community such methods have become known under the names “method of sieves,” “generalized series model” (see paper by Liang et al. in section 14.8), “finite series expansion methods” (see Censor’s work in the suggested reading list), and others.

In each of these regularization techniques, the choice of the regularization parameter—be it α , σ , k , or h —is crucial. For most problems in medical imaging, choices based on simulations, trial and error, and experience suffice. However, there exist mathematically justified methods for choosing these parameters. Some are completely deterministic, such as the discrepancy principle and the L -curve. Others, such as generalized cross-validation (GCV), are stochastic in nature.

14.4.5 Maximum Entropy

In the maximum entropy (ME) method one determines a non-negative solution of the (underdetermined) system 14.46 by maximizing a measure for the entropy of f , e.g., the Shannon entropy

$$H_S = - \int f(x) \log x \, dx$$

or a discrete version of it. Many algorithms have become known for maximizing entropy, most notably the MART algorithm in CT (see the paper by Herman and Lent listed in section 14.8).

For the nonlinear problem

$$A(f) = g \tag{14.55}$$

with a nonlinear operator A , the well-known Newton method reads

$$A'(f_k)(f_{k+1} - f_k) = g - A(f_k) . \tag{14.56}$$

Here, A' denotes the derivative of A . If equation 14.46 is ill posed, equation 14.56 has to be regularized, e.g., by the Tikhonov-Phillips method. Other iterative methods for solving equation 14.55 are the conjugate gradient algorithm and the Levenberg-Marquardt algorithm.

14.5 Sampling

All the reconstruction algorithms in medical imaging have to be implemented in a discrete setting. Questions of how densely and in which way the data have to be sampled in order to get a certain image quality are of paramount interest. We distinguish between sampling in real space and sampling in Fourier space.

14.5.1 Sampling in Real Space

Sampling in real space occurs, for example, in transmission CT. In principle it can be dealt with using classical sampling theory (such as is associated with the names Shannon, Petersen-Middleton, and Beurling). The two-dimensional case is fairly well understood. Assume that the reconstruction region is the circle with radius ρ , and that we want to reconstruct reliably functions that are essentially ω -band-limited, i.e., whose Fourier transform is negligible in some sense outside the circle of radius ω . According to the Shannon sampling theorem, this restriction on the Fourier transform corresponds to a spatial resolution of $2\pi/\omega$. One of the classical results of transmission CT is that one needs $\frac{1}{\pi}\omega^2\rho^2$ pieces of data to reconstruct such a function reliably. An appropriate sampling geometry is the interlaced parallel geometry suggested as early as 1978 by Cormack (whose 1978 paper is listed in section 14.8). Fan beam sampling requires only slightly more pieces of data (see Natterer's 1993 paper on fan beam sampling listed in the suggested reading). Algorithms are available that actually achieve the resolution $2\pi/\omega$ for these data sets (see, e.g., Faridani in section 14.8).

Not much is known about sampling in three dimensions, not even in fairly well developed fields such as helical scanning in transmission CT or in three-dimensional PET. It seems quite possible that the techniques used in two dimensions lead to substantial data reduction in three dimensions.

14.5.2 Sampling in Fourier Space

Sampling in Fourier space is quite different from sampling in real space and is much less well understood. The problem of recovering a function f from scattered values $\hat{f}(\xi_j)$ of its Fourier transform arises in many fields of medical imaging, e.g., in two- and three-dimensional transmission tomography and in MRI; compare to section 4.3. If the ξ_j are the points of a regular cartesian grid the problem can be easily solved by the fast Fourier transform (FFT). However, in two-dimensional transmission tomography the ξ_j are the points of a grid in polar coordinates. This occurs also in some instances of MRI. In other instances, the ξ_j are sitting on a spiral, or they are distributed in a non-symmetric way with respect to the origin. Many methods have been developed for dealing with this situation: the gridding method, a method based on the principle of the stationary phase, methods that estimate the phase of f based on small samples, localized polynomial approximation, autoregressive moving averages (ARMAs), projection on convex subsets (POCS), and maximum entropy (compare section 14.4 and see the paper by Liang et al. in section 14.8). Sometimes these methods work quite well, but clearly better methods are needed and are likely to exist.

14.6 Priors and Side Information

The regularization techniques mentioned in section 14.4 permit one to incorporate non-specific side information, such as smoothness and size. In many cases, quite specific side information is at hand, and much effort has been made to make use of such information. An early example is the theory of optimal recovery (Michelli and Rivlin in suggested reading list). An optimal recovery scheme S for the problem 14.46 with side information $f \in M$ is defined as the minimizer of the worst-case error

$$\sup\{\|f - Sg^\delta\| : f \in M, \|Af - g^\delta\| \leq \delta\}$$

among all operators $S : Y \rightarrow X$. While this approach looks natural to a mathematician, it obviously has not been of much use. Another natural approach is to minimize

$$\|Af - g^\delta\|$$

in M . This constraint optimization approach has been used with some success. But it is easy to construct examples in which ignoring the (correct!) side information $f \in M$ yields better results. The conclusion from these examples is that the obvious things have already been tried.

We now describe some more recent approaches, beginning with POCS. Assume that the set M is the intersection of convex subsets M_1, \dots, M_m , and put $M_0 = \{f : \|Af - g^\delta\| \leq \delta\}$. Then it is natural to define a sequence f_k of approximations to f by successively projecting on M_0, M_1, \dots, M_m . More specifically, let P_j , $j = 0, \dots, m$ be the orthogonal projections of X onto M_j . Then,

$$f^{k+1} = P_m \dots P_0 f^k, \quad k = 0, 1, \dots \quad (14.57)$$

The well-known algebraic reconstruction technique (ART) or Kaczmarz method, which was used in the first commercially available CT-scanner, is a special case of 14.57 with M_j the j -th hyperplane in the system 14.46.

Typical examples of the sets M_j are $M_1 = \{f : a \leq f \leq b\}$, $M_2 = \{f : f = 0 \text{ on some subset}\}$, and $M_3 = \{f : \|D^\beta f\| \leq \rho\}$ with a derivative of order β .

POCS permits the use of any side information that can be expressed by convex sets. Even though it has been used successfully for various problems, the mathematical theory is not well developed, and algorithms for POCS tend to be slow. Special methods have been developed for exploiting non-negativity (e.g., by Janssen; see section 14.8).

Rather than using deterministic side information that can be expressed by some subset M , one can also use stochastic side information. An example is the well-known Bayesian method. Here we think of f and g as families of random variables that are jointly normally distributed with mean values f_0 and g_0 and covariance operators F and G , respectively; i.e.,

$$F(x, x') = E[f(x) - f_0(x)][f(x') - f_0(x')], \quad (14.58)$$

with E the mathematical expectation, and correspondingly for G . Equation 14.46 is replaced by the linear stochastic model

$$Af = g + n, \quad (14.59)$$

where n is a family of random variables independent of f that is normally distributed with mean value 0 and covariance Σ . The Bayesian estimate for f given g is

$$f_B = E(f|g) = f_0 + FA^*(AFA^* + \Sigma)^{-1}(g - A_0f). \quad (14.60)$$

The similarity to equation 14.51 is obvious. In fact Bayesian methods very often yield results very similar to those of the Tikhonov-Phillips methods, but provide more flexibility in the modeling of the noise and the properties of f .

A plethora of numerical methods for computing f_B have been developed. Most of them are iterative. In the case of circular symmetry, direct methods can be used, too.

In medical imaging, the random variables are usually not normally distributed. Then, the Bayesian estimate is no longer a linear function of the data, and one has to resort to other techniques. One of them is expectation maximization (EM) already mentioned on page 95. In any case, Bayesian methods require a prior F , which models the interpixel correlation of the image f . Various choices for F can be found in the paper by Geman and McClure in the suggested reading list.

14.7 Research Opportunities

- Investigation of the trade-offs of stability versus resolution for the inverse problem for the Helmholtz equation, as applied to ultrasound and microwave imaging.

- Development of the mathematical theory and algorithms for inverse problems for the transport equation. There are applications not only to light tomography but also to scatter correction in CT and emission CT. Also needed are investigations of diffusion and diffusion-wave approximations.
- Development of numerical methods for parabolic inverse problems for application to light tomography (the diffusion approximation). Methods using adjoint fields look promising.
- Investigation of Gelfand-Levitan theory for multi-dimensional hyperbolic inverse problems. The one-dimensional Gelfand-Levitan theory is the backbone for one-dimensional inverse scattering; extension to three dimensions would solve the mathematical and numerical problems in ultrasound and microwave imaging.
- Development of a general-purpose algorithm for bilinear inverse problems. The inverse problems of medical imaging frequently have a bilinear structure, irrespective of the type of underlying equation (elliptic, parabolic, hyperbolic, or transport). A general-purpose algorithm for discretized problems is conceivable and would save a lot of work.
- Development of methods of scatter correction through transport models for transmission and emission CT. Preliminary results are already available. This challenge should be easier than the inverse problem in light tomography, at least for cases in which scatter is not too large.
- Creation of reconstruction algorithms for three-dimensional CT and efficient algorithms for cone-beam and helical scanning.
- Classification of three-dimensional scanning geometries according to the stability of the inversion problem.
- Development of reconstruction algorithms, possibly of Fourier type, for three-dimensional PET. Recent progress has been made by use of the stationary phase principle.
- Development of faster methods for computing maximum likelihood estimates with priors, more efficient iterative methods, and methods that exploit symmetries of the scanning geometries through efficient numerical algorithms such as the FFT.
- Investigation of preconditioning for nonlinear iterations such as expectation maximization (EM).
- Construction of good priors for SPECT and PET computations.
- Creation of mathematical attenuation corrections for emission CT, i.e., determination of the attenuation map from the emission data (without transmission measurements). Encouraging mathematical results for two source distributions are available, and simulations with templates have been performed.
- Creation of specialized regularization methods, particularly for regularization in time, which would have applications to magnetic and electrical source imaging.
- Reconstruction of functions under global shape information using templates. A typical application is the reconstruction of attenuation maps for emission CT and magnetic and electrical source imaging.
- Constraint reconstruction, i.e., reconstruction of a function from transmission or emission CT or MRI data with side conditions such as non-negativity. There are applications to limited-angle CT and to MRI scans with insufficient sampling.
- Reconstruction of a function from irregularly spaced samples of its Fourier transform, which has applications to CT and MRI.
- Removal of artifacts caused by opaque objects, such as hip joints in CT. Many papers are available, but there is no satisfactory solution.
- Reconstruction of a function in \mathbf{R}^3 from integrals over almost-planar surfaces. There are applications to MRI data collected with imperfect magnets.
- Reconstruction of vector fields. Are there ways to determine more than the curl? If not, what can be concluded from the curl? There are applications to Doppler imaging.

14.8 Suggested Reading

1. Aben, H.K. *Integrated Photoelasticity*, Valgus, Talin, 1975 (in Russian).
2. Anikonov, D.S., Uniqueness of simultaneous determination of two coefficients of the transport equation, *Sov. Math. Dokl.* **30** (1984), 149–151.
3. Anikonov, D.S., Prokhorov, I.V., and Kovtanyuk, E.E., Investigation of scattering and absorbing media by the methods of x-ray tomography, *J. Invest. Ill-Posed Probl.* **1** (1993), 259–281.
4. Arridge, S.R., Van der Zee, P., Cope, M., and Delpy, D.T., Reconstruction methods for infrared absorption imaging, *Proc. SPIE* **1431** (1991), 204–215.
5. Bondarenko, A., and Antyufeev, V., *X-ray Tomography in Scattering Media*, Technical report, Institute of Mathematics, Novosibirsk, Russia (1990).
6. Bronnikov, A.V., Degradation transform in tomography, *Pattern Recognition Letters* **15** (1994), 527–592.
7. Burridge, R., The Gelfand-Levitan, the Marchenko, and the Gopinath-Sondhi integral equations of inverse scattering theory, regarded in the context of inverse impulse-response problems, *Wave Motion* **2** (1980), 305–323.
8. Censor, Y., Finite series-expansion reconstruction methods, *Proc. IEEE* **71** (1983), 409–419.
9. Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, June 23–25, 1993, Snowbird, Utah, *Phys. Med. Biol.* **39** (1994).
10. Cormack, A.M., Sampling the Radon transform with beams of finite width, *Phys. Med. Biol.* **23** (1978), 1141–1148.
11. Faridani, A., Reconstructing from efficiently sampled data in parallel-beam computed tomography, in *Inverse Problems and Imaging*, G.F. Roach, ed., Pitman Res. Notes Math. **245**, John Wiley & Sons, New York, 68–102, 1991.
12. Geman, S., and McClure, D., Statistical methods for tomographic image reconstruction, *Bull. Int. Stat. Inst.* **LII** (1987), 5–21.
13. Gratton, E., et al., A novel approach to laser tomography, *Bioimaging* **1** (1993), 40–46.
14. Gutman, S., and Klibanov, M.V., Regularized quasi-Newton method for inverse scattering problems, *Math. Comput. Modeling* **18** (1993), 5–31.
15. Hämmäläinen, M.S., and Sarvas, J., Realistic conductivity model of the human head for interpretation of neuro-magnetic data, *IEEE Trans. Biomed. Eng.* **BME-36** (1989), 165–171.
16. Herman, G., *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, 1980.
17. Herman, G.T., and Lent, A., Iterative reconstruction algorithms, *Comput. Biol. Med.* **6** (1976), 273–294.
18. Hertle, A., The identification problem for the constant attenuation Radon transform, *Math. Z.* **197** (1988), 9–13.
19. Hinshaw, W.S., and Lent, A.H., An introduction to NMR imaging: From the Bloch equation to the imaging equation, *Proc. IEEE* **71** (1983), 338–350.
20. Isakov, V., Uniqueness and stability in multi-dimensional inverse problems, *Inverse Problems* **9** (1993), 579–617.
21. Janssen, A.J.E.M., Frequency-domain bounds for non-negative band-limited functions, *Philips J. Res.* **45** (1990), 325–366.
22. Kak, A.C., and Slaney, M., *Principles of Computerized Tomography Imaging*, IEEE Press, Los Alamidas, Calif., 1987.
23. Kleinman, R.E., and van den Berg, P.M., A modified gradient method for two-dimensional problems in tomography, *J. Comput. Appl. Math.* **42** (1992), 17–35.

24. Lavrentiev, M.M., Romanov, V.G., and Shishatskij, S.P., *Ill-Posed Problems of Mathematical Physics and Analysis*, Translations of Mathematical Monographs, vol. 64, American Mathematical Society, Providence, R.I., 1986.
25. Liang, Z.-P., Boada, F.E., Constable, R.T., Haacke, E.M., Lauterbur, P.C., and Smith, M.R., Constrained reconstruction methods in MR imaging, *Rev. Magn. Reson. Med.* **4** (1992), 67–185.
26. Louis, A.K., Medical imaging: State of the art and future development, *Inverse Problems* **8** (1992), 709–738.
27. Michelli, C.D., and Rivlin, T.J., eds., *Optimal Estimation in Approximation Theory*, Plenum Press, New York, 1977.
28. Natterer, F., Determination of tissue attenuation in emission tomography of optically dense media, *Inverse Problems* **9** (1993), 731–736.
29. Natterer, F., Sampling in Fan Beam Tomography, *SIAM J. Appl. Math.* **53** (1993), 358–380.
30. Natterer, F., *The Mathematics of Computerized Tomography*, John Wiley & Sons, New York, 1986.
31. Palamodov, V., An inversion method for attenuated x-ray transform in space, to appear in *Inverse Problems*.
32. Quinto, E.T., Cheney, M., and Kuchment, P., eds., *Tomography, Impedance Imaging, and Integral Geometry*, Proceedings of the AMS-SIAM Summer Seminar, June 7–18, 1993, Mount Holyoke College, Massachusetts, American Mathematical Society, Providence, R.I., 1993.
33. Romanov, V.G., Conditional stability estimates for the two-dimensional problem of restoring the right-hand side and absorption in the transport equation (in Russian), *Sib. Math. J.* **35** (1994), 1184–1201.
34. Sanchez, R., and McCormick, N.J., General solutions to inverse transport problems, *J. Math. Phys.* **22** (1981), 847–855.
35. Sharafutdinov, V.A., *Integral Geometry of Tensor Fields*, VSP, Utrecht, The Netherlands, 1994.
36. Shepp, L.A., and Vardi, Y., Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging* **1** (1982), 115–122.
37. Singer, J.R., Grünbaum, F.A., Kohn, P., and Zubelli, J.R., Image reconstruction of the interior of bodies that diffuse radiation, *Science* **248** (1990), 990–993.
38. Smith, K.T., Solmon, D.C., and Wagner, S.L., Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. Am. Math. Soc.* **83** (1977), 1227–1270.
39. Sparr, G., Stråklén, K., Lindström, K., and Persson, W., Doppler tomography for vector fields, *Inverse Problems* **11** (1995), 1051–1061.
40. Tretiak, O.J., and Metz, C., The exponential Radon transform, *SIAM J. Appl. Math.* **39** (1980), 341–354.

Index

- 3D PET data acquisition, 52
- 3D attenuation correction, 53
- 3D data acquisition in PET, 52
- 3D display, 32
- 3D display methods, 58
- 3D displays, 11, 12
- 3D reconstruction algorithms, 54
- 3D sampling, 104
- 3D visualization, 19, 80, 90

- acoustical holography, 56
- acoustical index of refraction, 56
- adaptive current tomograph (ACT), 66
- aliasing, 27
- amplifier design, 31
- analytical inversion, 53
- anatomical priors, 90
- angiographic imaging techniques, 27
- angiography, 22
- angioplasty, 79
- anisotropic diffusion, 36
- applied potential tomography (APT) system, 66
- artifacts, 10, 26
- attenuation, 8, 11, 42, 57
- attenuation coefficient distribution, 44
- attenuation coefficient distribution, non-uniform, 42, 45
- attenuation coefficient distribution, uniform, 45
- attenuation coefficients, 9
- attenuation compensation, 45
- attenuation distribution, 73, 94
- attenuation in SPECT, 42, 44, 45
- attenuation in ultrasound, 57
- avalanche photodiodes, 50

- back propagation in ultrasound, 58
- backprojection, 97
- Bayesian analysis, 24, 26
- Bayesian method, 26, 95, 105
- beam forming, 55, 56
- beam hardening, 92
- beam hardening artifacts, 9
- beam hardening effects, 10, 11
- beam propagation, 56
- BEM and FEM, 61
- biopsy, 78, 80
- Biot-Savart law, 100
- Bloch equation, 98
- Bloch equations, 16
- blood flow, 27, 30, 48
- blood flow imaging, 15
- blood oxygen level dependent (BOLD) effect, 30
- Born approximation, 95
- boundary element method, 61
- boundary element methods, 68
- brain SPECT, 42

- cardiac imaging, 6
- Cardiac SPECT, 42
- cardiac SPECT, 45
- cardiac strain, 30
- Cauchy problem, 61, 100
- change, 11, 16, 18, 20, 22, 32, 53, 58, 70, 73, 88
- chemical shift imaging, 33
- chest radiography, 3–5
- classical integral geometry, 99
- cluster analysis, 88
- coil design, 23, 35
- coil designs, 35
- coil size, 35
- collimator design, cone-beam, 46
- collimator design, 43
- collimator design, astigmatic, 46
- collimator design, cone-beam, 46
- collimator design, converging-hole, fan-beam, cone-beam, parallel-hole, 43
- collimator design, variable focal, 46
- collimator designs for SPECT, 46
- collimator designs in SPECT, 43
- collimator-detector response, 45
- collimator-detector system, 46
- Compton scatter, 53
- Compton scattering, 41–43, 49
- computational anatomy, 89
- computed tomography (CT), 8
- computer tomography (CT), 3
- computer-assisted visualization, 77
- conductivity distribution, 65
- cone beam errors, 12
- contrast agents, 74
- contrast agents, contrast-to-noise ratio, 35
- contrast resolution, 8
- convergence, 95
- converging-beam SPECT, 46
- convolution kernels, 12
- convolution methods, 44
- cryosurgery, 81, 82, 90
- CT, 72, 73, 78
- CT angiography, 13

- data acquisition, 12, 13, 17, 19, 20, 24–28, 31, 42, 43, 46, 51, 57, 60, 72, 87, 90
- data acquisition systems, 10
- data collection for optical methods, 72
- data correction, 10
- data corrections in PET, 51
- data display, 28
- data fitting, 89
- data fusion, 87
- data interpolation, 11
- data production with optical imaging, 72
- data reduction, 12, 52, 87, 90
- data reduction and display, 13
- data sampling, 53
- data visualization, 87
- detector array for MSI, 67
- detector arrays, 12
- detector efficiency in PET, 49
- detector elements, 10
- detector elements in PET, 49
- detector noise, 69
- detector ring based systems, 10
- detector scintillators, 10
- dielectric constant, changes in, 65
- diffraction integral, 56
- diffraction tomography, 96
- diffraction tomography for ultrasonics, 58
- diffusion, 35, 36
- diffusion coefficients, 29
- diffusion imaging, 21
- diffusion measurements, 29
- diffusion MRI, 29
- diffusion MRI in stroke, 36
- diffusion tensor imaging (DTI), 36
- digital subtraction angiography (DSA), 3
- discrete Fourier transform, 17
- display, 30, 36
- display methods, 57
- Doppler imaging, 106
- Doppler shift, 99
- Doppler tomography, 99
- dynamic 3D SPECT, 42
- dynamic magnetic resonance imaging, 15
- dynamic MRI, 24

- echo planar imaging (EPI), 20, 22, 29–31, 33, 36
- echo-planar imaging, 21, 24
- echo-planar imaging (EPI), 26, 27
- eddy current, 36
- eddy currents, 26, 32, 33
- eddy currents in MR, 21
- edge detection, 88
- EIT, 65, 101
- elastic warping, 89
- electrical activity of brain, 60
- electrical impedance tomography, 65, 97
- electrocardiography, 60
- electrocardiography (ECG), 60
- electroencephalography, 60

- electron beam systems, 12, 13
- electron beam techniques, 12
- electronics for EIT, 66
- emission CT, 94
- emission tomography, 94
- endocardial maps, 61
- endocardial potentials, 60
- endocardium, 61
- endoscopy, 77, 78
- energy resolution, 53, 54
- EPI, 79
- ESI, 100
- expectation-maximization algorithm, 51

- fan beam reconstruction algorithm, 9
- fan beam systems, 9, 10
- fan-beam systems, 11
- fast data acquisition in MRI, 81
- Fermi motion, 50
- FFT, 24, 26, 104
- field inhomogeneities, 25
- film subtraction angiography, 3
- filtered backprojection, 12, 93
- filtered backprojection algorithm, 51, 95
- filtered backprojection algorithm (3D), 52
- filtered backprojection of filtered projection, 44
- filtered-backprojection algorithm (3D), 53
- finite element calculations, 19
- finite element method, 61
- finite element methods, 68
- finite element models, 33, 56
- fitting algorithm, 89
- fMRI, 31, 32
- focused ultrasound, 82, 83
- forward models, 70
- forward problem, 61
- forward problems, 101
- Fourier optics, 56
- Fourier projection theorem, 44, 51
- Fourier reconstruction, 53
- Fourier space interpolation, 99
- Fourier transform, 17, 22, 25, 53, 96, 98, 99, 104
- functional magnetic resonance imaging (fMRI), 30
- functional MRI, 31, 32, 67
- fuzzy logic, 88

- Gelfand-Levitan procedure, 101
- Gibbs ringing, 24, 25
- Gibbs-Markov random field, 95
- global thresholding, 88
- gradient coil designs, 22
- gradient coils, 31, 32
- gradient coils for MR, 21
- gradient descent method, 22

- Hölder stability, 102
- half-value layer, 42
- handling of large wave numbers, 96
- heart arrhythmia, 61
- heat surgery, 81

- Helmholtz equation, 105
- high-speed CT, 11
- high-temperature superconductor magnets, 17
- high-temperature superconductors, 15
- histogramming, 88
- holography, 7
- Hounsfield numbers, 11
- hyperpolarized gases, 34

- ill-posedness, 26, 52, 61, 65, 73, 97, 100, 102
- image combination, image fusion, 20
- image contrast, 29
- image enhancement, 87
- image filtering, 24
- image fusion, 90
- image processing, 87, 90
- image processing methods, 28
- image reconstruction, 9, 11, 12, 15, 20, 24, 26, 42, 45, 49, 52, 60, 70
- image registration, 80, 81, 88–91
- image segmentation, 87–90
- image-guided therapy, 77
- imaging pulse sequences, 15
- immobilization during intervention, 80
- impedance tomography, 101
- instability, 97, 100–102
- instrument instability, 31
- instrumentation for ultrasonics, 55
- interpolation, 20
- interventional MRI, 29
- interventional radiology, 77
- inverse problem, 51, 61, 62, 67, 93–97, 100, 101, 105, 106
- inverse problems, 101
- inverse scattering problem for ultrasound, 58
- inversion algorithms for EIT, 65
- inversion problem, 73
- iterative algorithms, conjugate gradient (CG), 45
- iterative algorithms, expectation maximization (EM), 45
- iterative algorithms, Gauss-Seidel (GS), 45
- iterative reconstruction, 45, 46, 69
- iterative reconstruction algorithms, 45
- iterative reconstruction methods, 45

- k-space coverage, 24
- keyhole imaging, 26
- Klein-Nishina formula, 53

- Lagrange multipliers, 22
- LANDSAT, 88
- laser optical tomography, 72, 73
- layer stripping, 97
- Levenberg-Marquardt iterative method, 104
- Levenberg-Marquardt method, 101
- light tomography, 97
- limited-angle reconstruction, SPECT, 46
- linear attenuation coefficient, 8
- local gradient coil, 21
- local gradient coils, 21
- local RF coils, 21
- logarithmic stability, 102

- magnetic field for MRI, 17
- magnetic field gradients, 31
- magnetic resonance angiography (MRA), 33
- magnetic resonance imaging (MRI), 16
- magnetic resonance (MR), 15
- Magnetic source imaging (MSI), 67–69
- magnetocardiography (MCG), 67
- magnetoencephalography (MEG), 67
- magnets for MR, 17
- mammography, 3, 4
- maximum entropy, 45, 89
- maximum entropy method, 104
- maximum entropy methods, 26
- maximum likelihood, 45
- maximum likelihood method, 95
- measurement of oxygen, 30
- measuring electrical conductivity, 65
- Metz filter, 46
- microscopic imaging, 35
- minimally invasive surgery, 77
- Monte Carlo methods, 22
- motion artifacts, 10, 29
- motion detection, 88
- movement during data acquisition, 90
- MR flow imaging, 28
- MRI, 3, 73, 77–79, 81, 88, 97–99
- MRI and intervention, 79
- MRI angiogram, 90
- MRI magnets, 81
- MRI microscopy, 36
- MRS, 33
- MSI, 100
- multicoil arrays, 23
- multidimensional data, 87
- multiple coils, 20
- multiple detector SPECT systems, 42
- multiple receivers, 20

- neural networks, 26, 88
- Newton method, 101
- Newton's method, 97
- noise, 10, 31, 42, 45, 53, 61, 70, 73
- nuclear medicine imaging, 41, 43

- objective function, 62
- optical tomography, 74, 96, 97, 101

- paramagnetic contrast agents, 34
- parametric estimation techniques, 26
- partial volume coils, 21
- patient management, 77
- patient-based noise, 19
- patient-related artifacts, 10
- PCA combination, 20
- perturbation theory, 58
- PET, 67, 73, 94
- PET detector systems, 51

- PET detectors, 50, 53
- PET image reconstruction, 51
- PET scintillators, 49
- phase aberration, 56
- phase errors, 25
- phase mapping methods, 27, 28, 30
- phase-contrast MR imaging, 27
- phased array transducers, 56
- phased-array system, 56
- photoelectric interaction, 41
- photomultiplier tubes, 41, 49
- photon attenuation and scatter, 44
- physics-related artifacts, 10
- physiological noise, 32
- positron emission tomography (PET), 48–54
- precession frequency, 16
- prior information (use of), 70
- priors, 70, 95
- projection on convex subsets (POCS), 105
- pulsatile flow, 27
- pulse sequences, 79
- pulse-echo techniques, 57
- pulsed-field MRI, 18

- quantum efficiency, 10
- quantum noise, 10

- radiopharmaceuticals, 41, 42
- Radon inversion formula, 92
- Radon transform, 92, 98, 99
- Radon transform, attenuated, 44
- Radon transform, inverse, 44
- real-time imaging, 77
- reconstruction, 9, 10, 24, 43, 44, 51, 61, 69, 72
- reconstruction algorithms, 13
- reconstruction methods, 12, 26
- region growing, 88
- regularization, 100–103, 105
- regularization by Tikhonov-Phillips method, 104
- regularization methods, 62, 102
- reprojection techniques, 90
- resolution, 26, 28, 35, 41–44, 50, 52, 55, 57, 58, 65, 68, 69, 74, 96, 104
- resolution (spatial), 8, 12
- resolution in EIT, 65
- resolution limits in PET, 50
- resolution of MSI, temporal, 69
- resolution, ESI, 61
- restoration filter, 45
- RF coil, 17
- RF coil design, 19, 21, 22
- RF coil efficiency, 22
- RF coils, 19
- ring detector based systems, 9, 10
- ring-shaped artifacts, 11
- robot-assisted surgery, 77
- Rytov approximation, 58

- sampling, 12, 26, 56, 104
- sampling in Fourier space, 104

- scan times, 8
- scanner aperture, 53
- scatter, 93
- scatter compensation, 45
- scatter correction, 6, 53
- scatter correction in PET, 53
- scatter response function, 45
- scattering, 57, 73
- scintillation cameras, 43, 46
- scintillation crystals, 49
- scintillation detectors, 42
- segmentation, 34, 88
- segmentation, spectral techniques, 88
- semiconductor detectors for SPECT, silicon, germanium, cadmium zinc telluride, 43
- sensors for MSI, 67
- shape analysis, 88
- signal processing, 27, 35, 56, 57
- simplex, 89
- simplex method, 22
- simulated annealing, 22, 89
- Single-photon emission computed tomography (SPECT), 41–46
- singular value decomposition, 90
- SNR, 22, 24, 26, 27, 35, 52, 53, 69, 89
- spatial resolution, 8, 20, 22, 24–26, 41
- spatial resolution in SPECT, 42, 43
- spatial resolution of MSI, 69
- speckle, 56
- SPECT, 67, 73, 94
- SPECT detectors, 43
- SPECT image reconstruction, 42
- SPECT imaging, 42
- SPECT imaging process, 41
- SPECT instrumentation, 41, 42
- SPECT reconstruction, 44
- SPECT reconstruction methods, 44, 45
- SPECT reconstruction methods, iterative, 45
- SPECT sensitivity, 43
- SPECT system design—see SPECT instrumentation, 42
- spectroscopic imaging, 33, 34
- spiral CT, 8, 11–13, 78
- spiral scan methods, 27
- SQUID (superconducting quantum interference device), 67, 68
- SQUID detectors, 70
- stability, 26, 73, 89, 93, 96, 97, 101, 102
- stability estimate, 102
- stability of inv. problem, 95, 96
- stability of inverse problem, 97
- stability of magnetic fields, 17
- statistical criteria for SPECT reconstruction, minimum mean squares error (MMSE), weighted least squares (WLS), maximum entropy (ME), maximum likelihood (ML), maximum a posteriori (MAP), 45
- statistical edge finding, 88
- stochastic model, 97, 105

- stochastic models, 94
- superconducting coils, 35
- surface coils, 35
- surface gradient coils, 21
- surface morphology, 88
- surface rendering, 90
- surgical planning, 80, 90
- synthesis of parametric images, 89
- system electronics for ultrasound, 56
- system electronics in PET, 50
- system-related artifacts, 11

- temporal regularization, 62
- temporal resolution, 20
- tensor tomography, 99
- therapeutic intervention, 77
- thermal surgery, 81, 82
- thresholding, 88
- Tikhonov regularization, 62
- Tikhonov-Phillips method, 105
- time domain methods, 96
- time-of-flight information, 54
- tissue fiber orientation, 36
- tracking during surgery, 81
- transducer development, 55
- transmission CT, 92, 94, 104
- transmission tomography, 101
- transport equation, 96, 97
- transport operator, 101
- transport problem, 101
- treatment planning, 90

- ultrasonic imaging, 55
- ultrasound, 3, 77, 95, 101
- ultrasound imaging, 55, 56, 78, 81, 82
- ultrasound transducers, 56, 83
- use of priors with MSI, 70

- vector tomography, 99
- vectorcardiography (VCG), 60
- visualization, 90
- volume imaging, 32
- volume rendering, 90

- wavelets, 24, 26
- whole body gradient coils, 21
- whole-volume coils, 21
- Wiener filter, 46
- Wolff-Parkinson-White syndrome, 60

- x-ray CT, 97
- x-ray fluoroscopy, 77
- x-ray projection images, 3
- x-ray projection imaging, 3
- x-ray sources, monochromatic, 13
- x-ray sources, multiple monochromatic, 7
- x-ray tomography, 101
- x-ray tube, 9, 10
- x-ray tubes, 9, 11–13
- xeromammography, 5