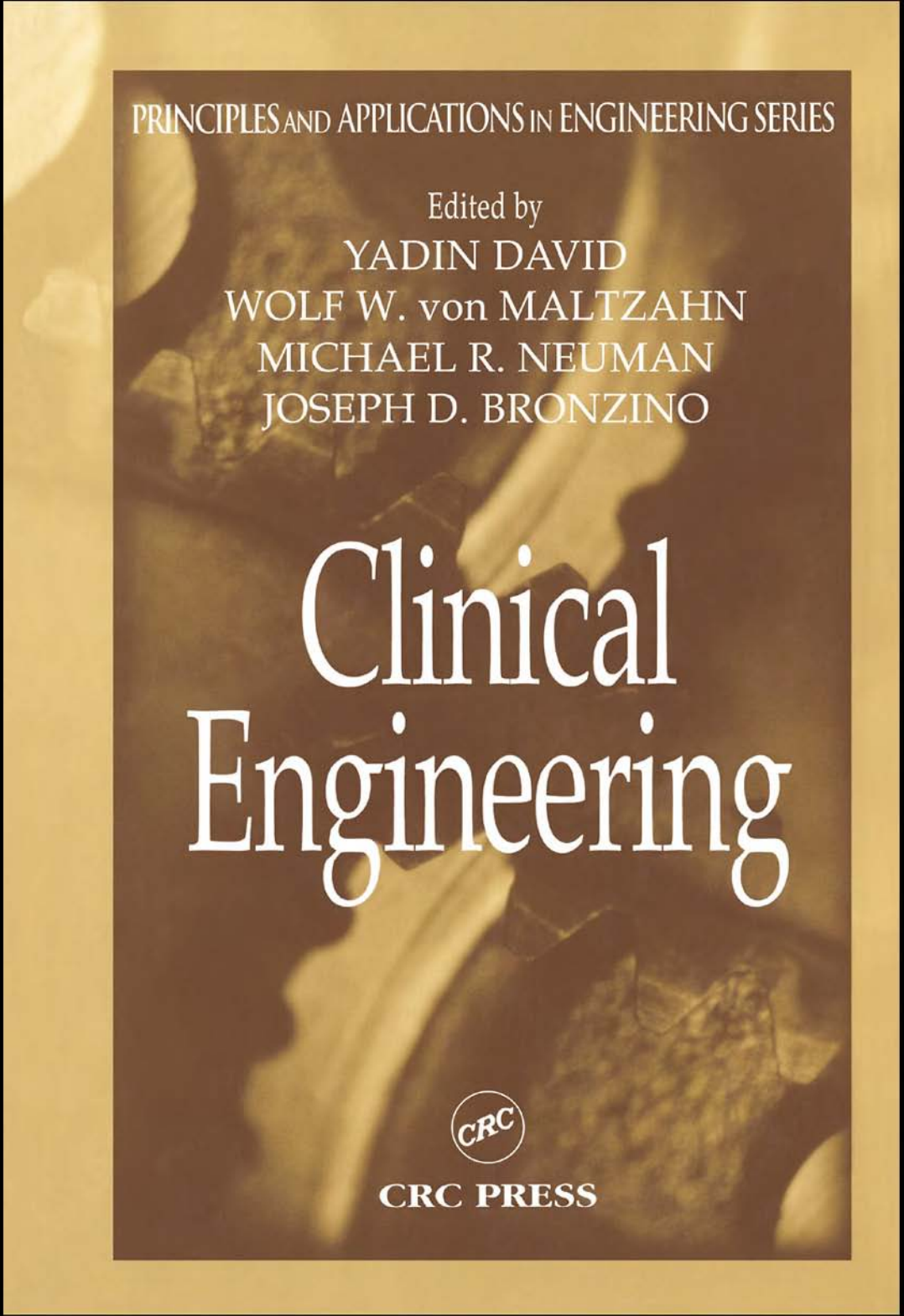


PRINCIPLES AND APPLICATIONS IN ENGINEERING SERIES

Clinical Engineering

PRINCIPLES AND APPLICATIONS IN ENGINEERING SERIES

Clinical Engineering

Edited by

YADIN DAVID

WOLF W.von MALTZAHN

MICHAEL R.NEUMAN

JOSEPH D.BRONZINO

CRC PRESS

Boca Raton London New York Washington, D.C.

This edition published in the Taylor & Francis e-Library, 2005.

The material in this book was first published in *The Biomedical Engineering Handbook.*

Library of Congress Cataloging-in-Publication Data
Clinical engineering/Yadin David…et al. p. cm. (Principles and applications in engineering) Includes biographical references and index. ISBN 0-8493-1813-0 (alk. paper) 1. Biomedical engineering. I. David, Yadin. II. Series. R856.543 2003 610'.28–dc21 2002041501

ISBN Master e-book ISBN

ISBN (Adobe e-Reader Format)
International Standard Book Number ISBN 0-8493-1813-0 (Print Edition)

Library of Congress Card Number 2002041501

Preface

By the turn of the 21st century, the field of clinical engineering, which is devoted to the application of engineering methods and technologies in the delivery of health care, has become well defined and widely accepted. Many hospitals have established centralized clinical engineering departments responsible for the management of the high technology devices and systems used in modern health care. As a result, clinical engineers are now called upon to provide the hospital administration with objective informed opinion regarding the assessment, acquisition, and utilization of medical technologies. Clinical engineers must be familiar with a wide range of medical devices, especially those related to the detection, display, and analysis of physiologic information.

The *Clinical Engineering Handbook* takes the sections most relevant to this important topic from the second edition of *The Biomedical Engineering Handbook*, published in 2000. This book opens with the section on Clinical Engineering edited by Dr. Yadin David, which provides an in-depth discussion of the methods used to manage deployment of medical technology, and integrates it appropriately with accepted clinical practices. The section focuses primarily on methodology for administering various critical engineering services and highlights the important roles that clinical engineers serve in many areas. It therefore emphasizes the importance of understanding the "bigger picture," which enables clinical engineers to have a much greater impact on their respective institutions and health care delivery in general.

Since familiarization with a wide range of medical instrumentation, especially those related to the detection, display, and analysis of physiologic function is critical to the clinical engineer, this book also covers both biomedical sensors and medical instruments. In the section on biosensors, Dr. Michael Neuman has provided an important introduction to this important field within biomedical engineering. More specifically, physical and chemical sensors, which include biopotential electrodes and optical sensors, are covered in considerable detail. The final section, edited by Dr. Wolf W.vonMaltzahn, is devoted to medical instruments and devices. This section provides an excellent introduction to the more traditional topics of bioinstrumentation, as well as some of the more recently developed instruments and devices.

Advisory Board

John Webster
University of Wisconsin
Madison, Wisconsin

Contributors

Joseph Adam
Premise Development Corporation
Hartford, Connecticut

Dennis D.Autio
Dybonics, Inc.
Portland, Oregon

Khosrow Behbehani
The University of Texas at Arlington and The University of Texas, Southwestern Medical Center at Dallas
Arlington, Texas

Bruce R.Bowman
EdenTec Corporation
Eden Prairie, Minnesota

Joseph D.Bronzino
Trinity College/Biomedical Engineering Alliance for Connecticut (BEACON)
Hartford, Connecticut

Mark E.Bruley
ECRI
Plymouth Meeting, Pennsylvania

Richard P.Buck
University of North Carolina
Chapel Hill, North Carolina

Robert D.Butterfield
IVAC Corporation
San Diego, California

Vivian H.Coates
ECRI
Plymouth Meeting, Pennsylvania

Yadin David
Texas Children’s Hospital
Houston, Texas

Gary Drzewiecki
Rutgers State University
Piscataway, New Jersey

Edwin G.Duffin
Medtronic, Inc.
Minneapolis, Minnesota

Jeffrey L.Eggleston
Valleylab, Inc.
Boulder, Colorado

Ross Flewelling
Nellcor Incorporation
Pleasant, California

Michael Forde
Medtronic, Inc.
Minneapolis, Minnesota

Leslie A.Geddes
Purdue University
West Lafayette, Indiana

Michael L.Gullikson
Texas Children's Hospital
Houston, Texas

Thomas M.Judd
Kaiser Permanente
Atlanta, Georgia

Millard M.Judy
Baylor Research Institute
Dallas, Texas

Chung-Chiun Liu
Electronics Design Center and Edison sensor Technology Center
Case Western Reserve University Cleveland, Ohio

Joseph P.McClain
Walter Reed Army Medical Center
Washington, D.C.

Yitzhak Mendelson
Worcester Polytechnic Institute
Worcester, Massachusetts

Robert L.Morris
Dybonics, Inc.
Portland, Oregon

Joachim H.Nagel
University of Stuttgart
Stuttgart, Germany

Michael R.Neuman
Joint Program in Biomedical Engineering
The University of Memphis and University of Tennessee
Memphis, Tennessee

Robert Patterson
University of Minnesota
Minneapolis, Minnesota

A.William Paulsen
Emory University
Atlanta, Georgia

P.Hunter Peckham
Case Western Reserve University and Veterans Affairs Medical Center
Cleveland, Ohio

Pat Ridgely
Medtronic, Inc.
Minneapolis, Minnesota

Richard L.Roa
Baylor University Medical Center
Dallas, Texas

Eric Rosow
Hartford Hospital and Premise Development Corporation
Hartford, Connecticut

Edward Schuck
EdenTec Corporation
Eden Prairie, Minnesota

Primoz Strojnik
Case Western Reserve University
Cleveland, Ohio

Willis A.Tacker
Purdue University
West Lafayette, Indiana

Wolf W.von Maltzahn
Whitaker Foundation
Washington, D.C.

Gregory I.Voss
IVAC Corporation
San Diego, California

Alvin Wald
Columbia University
New York, New York

Contents

Clinical Engineering

Yadin David
Texas Children's Hospital

OVER THE PAST 100 YEARS, the health care system's dependence on medical technology for the delivery of its services has grown continuously. To some extent, all professional care providers depend on technology, be it in the area of preventive medicine, diagnosis, therapeutic care, rehabilitation, administration, or health-related education and training. Medical technology enables practitioners to intervene through integrated interactions with their patients in a cost-effective, efficient, and safe manner. As a result, the field of clinical engineering has emerged as the discipline of biomedical engineering that fulfills the need to manage the deployment of medical technology and to integrate it appropriately with desired clinical practices.

The health care delivery system presents a very complex environment where facilities, equipment, materials, and a full range of human interventions are involved. It is in this clinical environment that patients of various ages and conditions, trained staff, and the wide variety of medical technology converge. This complex mix of interactions may lead to unacceptable risk when programs for moni-toring, controlling, improving, and educating all entities involved are not appropriately integrated by qualified professionals.

This section of clinical engineering focuses on the methodology for administering critical engineering services that vary from facilitation of innovation and technology transfer to the performance of technol-ogy assessment and operations support and on the management tools with which today's clinical engineer needs to be familiar. With increased awareness of the value obtained by these services, new career opportunities are created for clinical engineers.

In this section the authors have attempted to provide a description of the wide range of responsibilities clinical engineering professionals encounter. After presenting the evolution of the field of clinical engi-neering, Chapter 1 gives specific attention to the

primary function of clinical engineers. Chapter 2 describes technology management and assessment in considerable detail, using examples and case studies in a large medical center. Chapter 3 focuses on a particular technology management tool for assessing the medical equipment risk factor that can help clinical engineers effectively manage/prioritize the services to be provided to each piece of equipment under their control. To further assist clinical engineers in managing their equipment, Chapters 4 and 5 examine an establishment of program indicators that can lead to quality improvement. The clinical engineering program can be based in a single community hospital, in a teaching medical center, within a chain of hospitals, as part of a government agency, or as a shared service organization. Chapters 6 and 7 present a review of the standards and regulatory agencies of interest to clinical engineers while Chapter 8 introduces a new activity for clinical engineers—virtual instrumentation.

In addition to highlighting the important roles that clinical engineers serve in many areas, the section focuses on those areas of the clinical engineering field that enhance the understanding of the "bigger picture." With such an understanding, the participation in and contribution by clinical engineers to this enlarged scope can be fully realized. The adoption of the tools described here will enable clinical engineers to fulfill their new role in the evolving health care delivery system.

All the authors in this section recognize this opportunity and are here recognized for volunteering their talent and time so that others can excel as well.

1
Clinical Engineering: Evolution of a Discipline

Joseph D.Bronzino
Trinity College/Biomedical Engineering Alliance and Consortium (BEACON)

1.1 What Is a Clinical Engineer?

As discussed in the introduction to this handbook, biomedical engineers apply the concepts, knowledge, and techniques of virtually all engineering disciplines to solve specific problems in the biosphere, i.e., the realm of biology and medicine. When biomedical engineers work within a hospital or clinic, they are more properly called *clinical engineers*. But what exactly is the definition of the term *clinical engineer*? In recent years, a number of organizations, e.g., the American Heart Association [1986], the American Association of Medical Instrumentation [Goodman, 1989], the American College of Clinical Engineers [Bauld, 1991], and the *Journal of Clinical Engineering* [Pacela, 1991], have attempted to provide an appropriate definition for the term, *clinical engineer*. For the purposes of this handbook, a *clinical engineer* is an engineer who has graduated from an accredited academic program in engineering or who is licensed as a professional engineer or engineer-in-training and is engaged in the application of scientific and technological knowledge developed through engineering education and subsequent professional experience within the health care environment in support of clinical activities. Furthermore, the clinical environment is defined as that portion of the health care system in which patient care is delivered, and clinical activities include direct patient care, research, teaching, and public service activities intended to enhance patient care.

1.2 Evolution of Clinical Engineering

Engineers were first encouraged to enter the clinical scene during the late 1960s in response to concerns about patient safety as well as the rapid proliferation of clinical equipment, especially in academic medical centers. In the process, a new engineering discipline—clinical engineering—evolved to provide the technological support necessary to meet these new needs. During the 1970s, a major expansion of clinical engineering occurred, primarily due to the following events:

0-8493-1813-0/03/$0.00+$1.50

- The Veterans' Administration (VA), convinced that clinical engineers were vital to the overall operation of the VA hospital system, divided the country into biomedical engineering districts, with a chief biomedical engineer overseeing all engineering activities in the hospitals in that district.
- Throughout the United States, clinical engineering departments were established in most large medical centers and hospitals and in some smaller clinical facilities with at least 300 beds.
- Clinical engineers were hired in increasing numbers to help these facilities use existing technology and incorporate new technology.

Having entered the hospital environment, routine electrical safety inspections exposed the clinical engineer to all types of patient equipment that was not being maintained properly. It soon became obvious that electrical safety failures represented only a small part of the overall problem posed by the presence of medical equipment in the clinical environment. The equipment was neither totally understood nor properly maintained. Simple visual inspections often revealed broken knobs, frayed wires, and even evidence of liquid spills. Investigating further, it was found that many devices did not perform in accordance with manufacturers' specifications and were not maintained in accordance with manufacturers' recommendations. In short, electrical safety problems were only the tip of the iceberg. The entrance of clinical engineers into the hospital environment changed these conditions for the better. By the mid-1970s, complete performance inspections before and after use became the norm, and sensible inspection procedures were developed [Newhouse et al., 1989]. In the process, clinical engineering departments became the logical support center for all medical technologies and became responsible for all the biomedical instruments and systems used in hospitals, the training of medical personnel in equipment use and safety, and the design, selection, and use of technology to deliver safe and effective health care.

With increased involvement in many facets of hospital/clinic activities, clinical engineers now play a multifaceted role (Fig. 1.1). They must interface successfully with many "clients," including clinical staff, hospital administrators, regulatory agencies, etc., to ensure that the medical equipment within the hospital is used safely and effectively.

Today, hospitals that have established centralized clinical engineering departments to meet these responsibilities use clinical engineers to provide the hospital administration with an objective option of equipment function, purchase, application, overall system analysis, and preventive maintenance policies. Some hospital expertise, the hospital is in a far better position to make more effective use of its technological resources [Bronzino, 1986, 1992]. By providing health professionals with the needed assurance of safety, reliability, and efficiency in using new and innovative equipment, clinical engineers can readily identify poor-quality and ineffective equipment, thereby resulting in faster, more appropriate utilization of new medical equipment.

Typical pursuits of clinical engineers, therefore, include

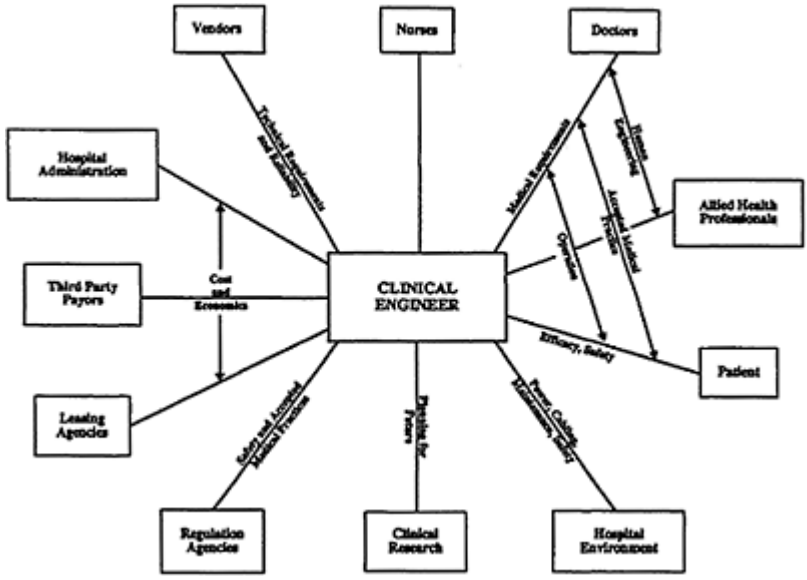


FIGURE 1.1 Diagram illustrating the range of interactions of a clinical engineer.

administrators have learned that with the in-house availability of such talent and

- Supervision of a hospital clinical engineering department that includes clinical engineers and biomedical equipment technicians (BMETs)
- Prepurchase evaluation and planning for new medical technology
- Design, modification, or repair of sophisticated medical instruments or systems
- Cost-effective management of a medical equipment calibration and repair service
- Supervision of the safety and performance testing of medical equipment performed by BMETs
- Inspection of all incoming equipment (i.e., both new and returning repairs)
- Establishment of performance benchmarks for all equipment
- Medical equipment inventory control
- Coordination of outside engineering and technical services performed by vendors
- Training of medical personnel in the safe and effective use of medical devices and systems
- Clinical applications engineering, such as custom modification of medical devices for clinical research, evaluation of new noninvasive monitoring systems, etc.
- Biomedical computer support
- Input to the design of clinical facilities where medical technology is used, e.g., operating rooms (ORs), intensive care units, etc.
- Development and implementation of documentation protocols required by external accreditation and licensing agencies.

Clinical engineers thus provide extensive engineering services for the clinical staff and, in recent years, have been increasingly accepted as valuable team members by physicians, nurses, and other clinical professionals. Furthermore, the acceptance of clinical engineers in the hospital setting has led to different types of engineering-medicine interactions, which in turn have improved health care delivery.

1.3 Hospital Organization and the Role of Clinical Engineering

In the hospital, management organization has evolved into a diffuse authority structure that is commonly referred to as the *triad model*. The three primary components are the governing board (trustees), hospital administration (CEO and administrative staff), and the medical staff organization [Bronzino and Hayes, 1988]. The role of the governing board and the chief executive officer are briefly discussed below to provide some insight regarding their individual responsibilities and their interrelationship.

Governing Board (Trustees)

The Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) summarizes the major duties of the governing board as "adopting by-laws in accordance with its legal accountability and its responsibility to the patient." The governing body, therefore, requires both medical and paramedical departments to monitor and evaluate the quality of patient care, which is a critical success factor in hospitals today.

To meet this goal, the governing board essentially is responsible for establishing the mission statement and defining the specific goals and objectives that the institution must satisfy. Therefore, the trustees are involved in the following functions:

- Establishing the policies of the institution
- Providing equipment and facilities to conduct patient care
- Ensuring that proper professional standards are defined and maintained (i.e., providing quality assurance)
- Coordinating professional interests with administrative, financial, and community needs
- Providing adequate financing by securing sufficient income and managing the control of expenditures
- Providing a safe environment
- Selecting qualified administrators, medical staff, and other professionals to manage the hospital

In practice, the trustees select a hospital chief administrator who develops a plan of action that is in concert with the overall goals of the institution.

Hospital Administration

The hospital administrator, the chief executive officer of the medical enterprise, has a function similar to that of the chief executive officer of any corporation. The administrator represents the governing board in carrying out the day-to-day operations to

reflect the broad policy formulated by the trustees. The duties of the administrator are summarized as follows:

- Preparing a plan for accomplishing the institutional objectives, as approved by the board
- Selecting medical chiefs and department directors to set standards in their respective fields
- Submitting for board approval an annual budget reflecting both expenditures and income projections
- Maintaining all physical properties (plant and equipment) in safe operating condition
- Representing the hospital in its relationships with the community and health agencies
- Submitting to the board annual reports that describe the nature and volume of the services delivered during the past year, including appropriate financial data and any special reports that may be requested by the board

In addition to these administrative responsibilities, the chief administrator is charged with controlling cost, complying with a multitude of governmental regulations, and ensuring that the hospital conforms to professional norms, which include guidelines for the care and safety of patients.

1.4 Clinical Engineering Programs

In many hospitals, administrators have established clinical engineering departments to manage effectively all the technological resources, especially those relating to medical equipment, that are necessary for providing patient care. The primary objective of these departments is to provide a broad-based engineering program that addresses all aspects of medical instrumentation and systems support.

Figure 1.2 illustrates the organizational chart of the medical support services division of a typical major medical facility. Note that within this organizational structure, the director of clinical engineering reports directly to the vice-president of medical support services. This administrative relationship is extremely important because it recognizes the important role clinical engineering departments play in delivering quality care. It should be noted, however, that in other common organizational structures, clinical engineering services may fall under the category of “facilities,” “materials management,” or even just “support services.” Clinical engineers also can work directly with clinical departments, thereby bypassing much of the hospital hierarchy. In this situation, clinical departments can offer the clinical engineer both the chance for intense specialization and, at the same time, the opportunity to develop personal relationships with specific clinicians based on mutual concerns and interests [Wald, 1989].

Once the hospital administration appoints a qualified individual as director of clinical engineering, the person usually functions at the department-head level in the organizational structure of the institution

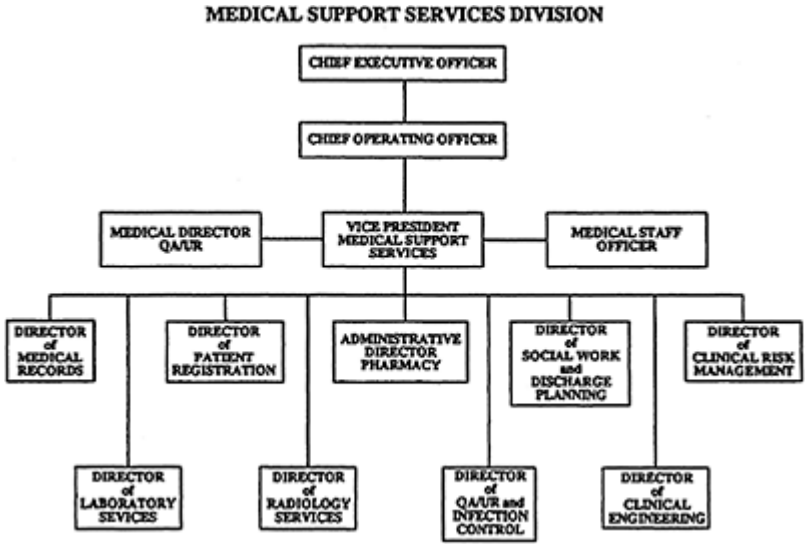


FIGURE 1.2 Organizational chart of medical support services division for a typical major medical facility. This organizational structure points out the critical interrelationship between the clinical engineering department and the other primary services provided by the medical facility.

and is provided with sufficient authority and resources to perform the duties efficiently and in accordance with professional norms. To understand the extent of these duties, consider the job title for "clinical engineering director" as defined by the World Health Organization [Issakov et al, 1990].

General Statement. The clinical engineering director, by his or her education and experience, acts as a manger and technical director of the clinical engineering department. The individual designs and directs the design of equipment modifications that may correct design deficiencies or enhance the clinical performance of medical equipment. The individual also may supervise the implementation of those design modifications. The education and experience that the director possesses enables him or her to analyze complex medical or laboratory equipment for purposes of defining corrective maintenance and developing appropriate preventive maintenance or performance-assurance protocols. The clinical engineering director works with nursing and medical staff to analyze new medical equipment needs and participates in both the prepurchase planning process and the incoming testing process. The individual also participates in the equipment management process through involvement in the system development, implementation, maintenance, and modification processes.

Duties and Responsibilities. The director of clinical engineering has a wide range of duties and responsibilities. For example, this individual

- Works with medical and nursing staff in the development of technical and performance specifications for equipment requirements in the medical mission.
- Once equipment is specified and the purchase order developed, generates appropriate testing of the new equipment.
- Does complete performance analysis on complex medical or laboratory equipment and summarizes results in brief, concise, easy-to-understand terms for the purposes of recommending corrective action or for developing appropriate preventive maintenance and performance assurance protocols.
- Designs and implements modifications that permit enhanced operational capability. May supervise the maintenance or modification as it is performed by others.
- Must know the relevant codes and standards related to the hospital environment and the performance assurance activities. (Examples in the United States are NFPA 99, UL 544, and JCAHO, and internationally, IEC-TC 62.)
- Is responsible for obtaining the engineering specifications (systems definitions) for systems that are considered unusual or one-of-a-kind and are not commercially available.
- Supervises in-service maintenance technicians as they work on codes and standards and on preventive maintenance, performance assurance, corrective maintenance, and modification of new and existing patient care and laboratory equipment.
- Supervises parts and supply purchase activities and develops program policies and procedures for same.
- Sets departmental goals, develops budgets and policy, prepares and analyzes management reports to monitor department activity, and manages and organizes the department to implement them.
- Teaches measurement, calibration, and standardization techniques that promote optimal performance.
- In equipment-related duties, works closely with maintenance and medical personnel. Communicates orally and in writing with medical, maintenance, and administrative professionals. Develops written procedures and recommendations for administrative and technical personnel.

Minimum Qualifications. A bachelor's degree (4 years) in an electrical or electronics program or the equivalent is required (preferably with a clinical or biomedical adjunct). A master's degree is desirable. A minimum of 3 years' experience as a clinical engineer and 2 years in a progressively responsible supervisory capacity is needed. Additional qualifications are as follows:

- Must have some business knowledge and management skills that enable him or her to participate in budgeting, cost accounting, personnel management, behavioral counseling, job description development, and interviewing for hiring or firing purposes. Knowledge and experience in the use of microcomputers are desirable.
- Must be able to use conventional electronic trouble-shooting instruments such as multimeters, function generators, oscillators, and oscilloscopes. Should be able to use conventional machine shop equipment such as drill presses, grinders, belt sanders, brakes, and standard hand tools.

- Must possess or be able to acquire knowledge of the techniques, theories, and characteristics of materials, drafting, and fabrication techniques in conjunction with chemistry, anatomy, physiology, optics, mechanics, and hospital procedures.
- Clinical engineering certification or professional engineering registration is required.

Major Functions of a Clinical Engineering Department

It should be clear from the preceding job description that clinical engineers are first and foremost engineering professionals. However, as a result of the wide-ranging scope of interrelationships within the medical setting, the duties and responsibilities of clinical engineering directors are extremely diversified. Yet a common thread is provided by the very nature of the technology they manage. Directors of clinical engineering departments are usually involved in the following core functions:

Technology Management. Developing, implementing, and directing equipment management programs. Specific tasks include accepting and installing new equipment, establishing preventive maintenance and repair programs, and managing the inventory of medical instrumentation. Issues such as cost-effective use and quality assurance are integral parts of any technology management program. The director advises the administrator of the budgetary, personnel, space, and test equipment requirements necessary to support this equipment management program.

Risk Management. Evaluating and taking appropriate action on incidents attributed to equipment malfunctions or misuse. For example, the clinical engineering director is responsible for summarizing the technological significance of each incident and documenting the findings of the investigation. He or she then submits a report to the appropriate hospital authority and, according to the Safe Medical Devices Act of 1990, to the device manufacturer, the Food and Drug Administration (FDA), or both.

Technology Assessment. Evaluating and selecting new equipment. The director must be proactive in the evaluation of new requests for capital equipment expenditures, providing hospital administrators and clinical staff with an in-depth appraisal of the benefits/advantages of candidate equipment. Furthermore, the process of technology assessment for all equipment used in the hospital should be an ongoing activity.

Facilities Design and Project Management. Assisting in the design of new or renovated clinical facilities that house specific medical technologies. This includes operating rooms, imaging facilities, and radiology treatment centers.

Training. Establish and deliver instructional modules for clinical engineering staff as well as clinical staff on the operation of medical equipment.

In the future, it is anticipated that clinical engineering departments will provide assistance in the application and management of many other technologies that support patient care, including computer support, telecommunications, and facilities operations.

Defining Terms

JCAHO, Joint Commission on the Accreditation of Healthcare Organizations: Accrediting body responsible for checking hospital compliance with approved rules and regulations regarding the delivery of health care.

Technology assessment: Involves an evaluation of the safety, efficiency, and cost effectiveness, as well as consideration of the social, legal, and ethical effects, of medical technology.

References

AHA, 1986. American Hospital Association Resource Center. *Hospital Administration Terminology,* 2nd ed. Washington, American Hospital Publishing.

Bauld TJ. 1991. The definition of a clinical engineer. *J Clin Eng* 16:403.

Bronzino JD. 1986. *Biomedical Engineering and Instrumentation: Basic Concepts and Applications.* Boston, PWS Publishing.

Bronzino JD. 1992. *Management of Medical Technology: A Primer for Clinical Engineers.* Boston, Butterworth.

Goodman G. 1989. The profession of clinical engineering. *J Clin Eng* 14:27.

ICC. 1991. *International Certification Commission's Definition of a Clinical Engineer, International Certification Commission Fact Sheet.* Arlington, Va, ICC.

Issakov A, Mallouppas A, McKie J. 1990. Manpower development for a healthcare technical service. Report of the World Health Organization, WHO/SHS/NHP/90.4.

Newhouse VL, Bell DS, Tackel IS, et al. 1989. The future of clinical engineering in the 1990s. *J Clin Eng* 14:417.

Pacela A. 1990. *Bioengineering Education Directory.* Brea, Calf, Quest Publishing.

Wald A. 1989. Clinical engineering in clinical departments: A different point of view. *Biomed Instr Technol* 23:58.

Further Information

Bronzino JD. 1992. *Management of Medical Technology: A Primer for Clinical Engineers.* Boston, Butterworth.

Journals: *Journal of Clinical Engineering, Journal of Medical Engineering and Physics, Biomedical Instrumentation and Technology.*

2
Management and Assessment of Medical Technology

Yadin David
Texas Children's Hospital

Thomas M.Judd
Kaiser Permanente

As medical technology continues to evolve, so does its impact on patient outcome, hospital operations, and financial efficiency. The ability to plan for this evolution and its subsequent implications has become a major challenge in most decisions of health care organizations and their related industries. Therefore, there is a need to adequately plan for and apply those management tools that optimize the deployment of medical technology and the facilities that house it. Successful management of the technology and facilities will ensure a good match between the needs and the capabilities of staff and technology, respectively. While different types and sizes of hospitals will consider various strategies of actions, they all share the need to manage efficient utilization of their limited resources and its monitoring. Technology is one of these resources, and while it is frequently cited as the culprit behind cost increases, the well-managed technology program contribute to a significant containment of the cost of providing quality patient care. Clinical engineer's skills and expertise are needed to facilitate the adoption of an objective methodology for implantation of a program that will match the hospital's needs and operational conditions.

Whereas both the knowledge and practice patterns of management in general are well organized in today's literature, the management of the health care delivery system and that of medical technology in the clinical environment has not yet reached that same high level. However, as we begin to understand the relationship between the methods and information that guide the decision-making processes regarding the management of medical technology that are being deployed in this highly complex environment, the role of the qualified clinical engineer becomes more valuable. This is achieved by reformulating the technology management process, which starts with the strategic planning process, continues with the technology assessment process, leads to the equipment planning and procurement processes, and finally ends with the assets management process. Definition of terms used in this chapter are provided at the end of the chapter.

2.1 The Health Care Delivery System

Societal demands on the health care delivery system revolve around cost, technology, and expectations. To respond effectively, the delivery system must identify its goals, select and define its priorities, and then wisely allocate its limited resources. For most organizations, this means that they must acquire only appropriate technologies and

0-8493-1813-0/03/$0.00+$1.50

manage what they have already more effectively. To improve performance and reduce costs, the delivery system must recognize and respond to the key dynamics in which it operates, must shape and mold its planing efforts around several existing health care trends and directions, and must respond proactively and positively to the pressures of its environment. These issues and the technology manager's response are outlined here: (1) technology's positive impact on care quality and effectiveness, (2) an unacceptable rise in national spending for health care services, (3) a changing mix of how Americans are insured for health care, (4) increases in health insurance premiums for which appropriate technology application is a strong limiting factor, (5) a changing mix of health care services and settings in which care is delivered, and (6) growing pressures related to technology for hospital capital spending and budgets.

Major Health Care Trends and Directions

The major trends and directions in health care include (1) changing location and design of treatment areas, (2) evolving benefits, coverages, and choices, (3) extreme pressures to manage costs, (4) treating of more acutely ill older patients and the prematurely born, (5) changing job structures and demand for skilled labor, (6) the need to maintain a strong cash flow to support construction, equipment, and information system developments, (7) increased competition on all sides, (8) requirement for information systems that effectively integrate clinical and business issues, (9) changing reimbursement policies that reduce new purchases and lead to the expectation for extended equipment life cycles, (10) internal technology planning and management programs to guide decision making, (11) technology planning teams to coordinate adsorption of new and replacement technologies, as well as to suggest delivery system changes, and (12) equipment maintenance costs that are emerging as a significant expense item under great administrative scrutiny.

System Pressures

System pressures include (1) society's expectations—highest quality care at the lowest reasonable price, where quality is a function of personnel, facilities, technology, and clinical procedures offered; (2) economic conditions—driven often by reimbursement criteria; (3) legal-pressures—resulting primarily from malpractice issues and dealing with rule-intensive "government" clients; (4) regulatory—multistate delivery systems with increased management complexity, or heavily regulated medical device industries facing free-market competition, or hospitals facing the Safe Medical Devices Act reporting requirements and credentialling requirements; (5) ethics—deciding who gets care and when; and (6) technology pressures—organizations having enough capabilities to meet community needs and to compete successfully in their marketplaces.

The Technology Manager's Responsibility

Technology mangers should (1) become deeply involved and committed to technology planning and management programs in their system, often involving the need for greater personal responsibilities and expanded credentials; (2) understand how the factors above

impact their organization and how technology can be used to improve outcomes, reduce costs, and improve quality of life for patients, (3) educate other health care professionals about how to demonstrate the value of individual technologies through involving financial, engineering, quality of care, and management perspective, and (4) assemble a team of caregivers with sufficient broad clinical expertise and administrators with planning and financial expertise to contribute their knowledge to the assessment process [16].

2.2 Strategic Technology Planning

2.2.1 Strategic Planning Process

Leading health care organizations have begun to combine strategic technology planning with other technology management activities in programs that effectively integrate new technologies with their existing technology base. This has resulted in high-quality care at a reasonable cost. Among those who have been its leading catalysts, ECRI (formerly the Emergency Care Research Institute) is known for articulating this program [4] and encouraging its proliferation initially among regional health care systems and now for single or multihospital systems as well [5]. Key components of the program include clinical strategic-planning, technology strategic planning, technology assessment, interaction with capital budgeting, acquisition and deployment, resource (or equipment assets) management, and monitoring and evaluation. A proper technology strategic plan is derived from and supports a well-defined clinical strategic plan [15].

Clinical and Technology Strategic Plan

Usually considered long-range and continually evolving, a clinical strategic plan is updated annually. For a given year, the program begins when key hospital participants, through the strategic planning process, assess what clinical services the hospital should be offering in its referral area. They take into account health care trends, demographic and market share data, and space and facilities plans. They analyze their facility's strengths and weaknesses, goals and objectives, competition, and existing technology base. The outcome of this process is a clinical strategic plan that establishes the organization's vision for the year and referral area needs and the hospital's objectives in meeting them.

It is not possible to adequately complete a clinical strategic plan without engaging in the process of strategic technology planning. A key role for technology managers is to assist their organizations throughout the combined clinical and technology strategic planning processes by matching available technical capabilities, both existing and new, with clinical requirements. To accomplish this, technology managers must understand why their institution's values and mission are set as they are, pursue their institution's strategic plans through that knowledge, and plan in a way that effectively allocates limited resources. Although a technology manager may not be assigned to develop an institution's overall strategic plan, he or she must understand and believe it in order to offer good input for hospital management. In providing this input, a technology manager

should determine a plan for evaluating the present state of the hospital's technological deployment, assist in providing a review of emerging technological innovations and their possible impact on the hospital, articulate justifications and provisions for adoption of new technologies or enhancement of existing ones, visit research facilities and laboratories and exhibit areas at major medical and scientific meetings to view new technologies, and be familiar with the institution and its equipment users' abilities to assimilate new technology.

The past decade has shown a trend toward increased legislation in support of more federal regulations in health care. These and other pressures will require that additional or replacement medical technology be well anticipated and justified. As a rationale for technology adoption, the Texas Children's Hospital focuses on the issues of clinical necessity, management support, and market preference, Addressing the issue of clinical necessity, the hospital considers the technology's comparison against medical standard of care, its impact on the level of care and quality of life, its improvement on intervention's accuracy and/or safety, its impact on the rate of recovery, the needs or desires of the community, and the change in service volume or focus. On the issue of management support, the hospital estimates if the technology will create a more effective care plan and decision-making process, improve operational efficiency in the current service programs, decrease liability exposure, increase compliance with regulations, reduce workload and dependence on user skill level, ameliorate departmental support, or enhance clinical proficiency. Weighting the issue of market preference, the hospital contemplates if it will improve access to care, increase customer convenience and/or satisfaction, enhance the organization's image and market share, decrease the cost of adoption and ownership, or provide a return on its investment.

Technology Strategic Planning Process

When the annual clinical strategic planning process has started and hospital leaders have begun to analyze or reaffirm what clinical services they want to offer to the community, the hospital can then conduct efficient technology strategic planning. Key elements of this planning involve (1) performing an initial audit of existing technologies, (2) conducting a technology assessment for new and emerging technologies for fit with current or desired clinical services, (3) planning for replacement and selection of new technologies, (4) setting priorities for technology acquisition, and (5) developing processes to implement equipment acquisition and monitor ongoing utilization. "Increasingly, hospitals are designating a senior manager (e.g., an administrator, the director of planning, the director of clinical engineering) to take the responsibility for technology assessment and planning. That person should have the primary responsibility for developing the strategic technology plan with the help of key physicians, department managers, and senior executives" [4].

Hospitals can form a medical technology advisory committee (MTAC), overseen by the designated senior manager and consisting of the types of members mentioned above, to conduct the strategic technology planning process and to annually recommend technology priorities to the hospital strategic planning committee and capital budget committee. It is especially important to involve physicians and nurses in this process.

In the initial technology audit, each major clinical service or product line must be analyzed to determine how well the existing technology base supports it. The audit can be conducted along service lines (radiology, cardiology, surgery) or technology function (e.g., imaging, therapeutic, diagnostic) by a team of designated physicians, department heads, and technology managers. The team should begin by developing a complete hospital-wide assets inventory, including the quantity and quality of equipment. The team should compare the existing technology base against known and evolving standards-of-care information, patient outcome data, and known equipment problems. Next, the team should collect and examine information on technology utilization to assess its appropriate use, the opportunities for improvement, and the risk level. After reviewing the technology users' education needs as they relate to the application and servicing of medical equipment, the team should credential users for competence in the application of new technologies. Also, the auditing team should keep up with published clinical protocols and practice guidelines using available health care standards directories and utilize clinical outcome data for quality-assurance and risk-management program feedback [6].

While it is not expected that every hospital has all the required expertise in-house to conduct the initial technology audit or ongoing technology assessment, the execution of this planning process is sufficiently critical for a hospital's success that outside expertise should be obtained when necessary. The audit allows for the gathering of information about the status of the existing technology base and enhances the capability of the medical technology advisory committee to assess the impact of new and emerging technologies on their major clinical services.

All the information collected from the technology audit results and technology assessments is used in developing budget strategies. Budgeting is part of strategic technology planning in that a 2-to 5-year long-range capital spending plan should be created. This is in addition to the annual capital budget preparation that takes into account 1 year at a time. The MTAC, as able and appropriate, provides key information regarding capital budget requests and makes recommendations to the capital budget committee each year. The MTAC recommends priorities for replacement as well as new and emerging technologies that over a period of several years guides that acquisition that provides the desired service developments or enhancements. Priorities are recommended on the basis of need, risk, cost (acquisition, operational and maintenance), utilization, and fit with the clinical strategic plan.

2.3 Technology Assessment

As medical technology continues to evolve, so does its impact on patient outcome, hospital operations, and financial resources. The ability to manage this evolution and its subsequent implications has become a major challenge for all health care organizations. Successful management of technology will ensure a good match between needs and capabilities and between staff and technology. To be successful, an ongoing technology assessment process must be an integral part of an ongoing technology planning and management program at the hospital, addressing the needs of the patient, the user, and the support team. This facilitates better equipment planning and utilization of the hospital's resources. The manager who is knowledgeable about his or her organization's culture,

equipment users' needs, the environment within which equipment will be applied, equipment engineering, and emerging technological capabilities will be successful in proficiently implementing and managing technological changes [7].

It is in the technology assessment process that the clinical engineering/technology manager professional needs to wear two hats: that of manager and that of engineer. This is a unique position, requiring expertise and detailed preparation, that allows one to be a key leader and contributor to the decision-making process of the medical technology advisory committee (MTAC).

The MTAC uses an *ad hoc* team approach to conduct technology assessment of selected services and technologies throughout the year. The *ad hoc* teams may incorporate representatives of equipment users, equipment service providers, physicians, purchasing agents, reimbursement mangers, representatives of administration, and other members from the institution as applicable.

Prerequisites for Technology Assessment

Medical technology is a major strategic factor in positioning and creating a positive community perception of the hospital. Exciting new biomedical devices and systems are continually being introduced. And they are introduced at a time when the pressure on hospitals to contain expenditures is mounting. Therefore, forecasting the deployment of medical technology and the capacity to continually evaluate its impact on the hospital require that the hospital be willing to provide the support for such a program. (*Note:* Many organizations are aware of the principle that an in-house "champion" is needed in order to provide for the leadership that continually and objectively plans ahead. The champion and the program being "championed" may use additional in-house or independent expertise as needed. To get focused attention on the technology assessment function and this program in larger, academically affiliated and government hospitals, the position of a chief technology officer is being created.) Traditionally, executives rely on their staff to produce objective analyses of the hospital's technological needs. Without such analyses, executives may approve purchasing decisions of sophisticated biomedical equipment only to discover later that some needs or expected features were not included with this installation, that those features are not yet approved for delivery, or that the installation has not been adequately planned.

Many hospitals perform technology assessment activities to project needs for new assets and to better manage existing assets. Because the task is complex, an interdisciplinary approach and a cooperative attitude among the assessment team leadership is required. The ability to integrate information from disciplines such as clinical, technical, financial, administrative, and facility in a timely and objective manner is critical to the success of the assessment. This chapter emphasizes how technology assessment fits within a technology planning and management program and recognizes the importance of corporate skills forecasting medical equipment changes and determining the impact of changes on the hospital's market position. Within the technology planning and management program, the focus on capital assets management of medical equipment should not lead to the exclusion of accessories, supplies, and the disposables also required.

Medical equipment has a life cycle that can be identified as (1) the innovation phase, which includes the concept, basic and applies research, and development, and (2) the adoption phase, which begins with the clinical studies, through diffusion, and then widespread use. These phases are different from each other in the scope of professional skills involved, their impact on patient care, compliance with regulatory requirements, and the extent of the required operational support. In evaluating the applicability of a device or a system for use in the hospital, it is important to note in which phase of its life cycle the equipment currently resides.

Technology Assessment Process

More and more hospitals are faced with the difficult phenomenon of a capital equipment requests list that is much larger than the capital budget allocation. The most difficult decision, then, is the one that matches clinical needs with the financial capability. In doing so, the following questions are often raised: How do we avoid costly technology mistakes? How do we wisely target capital dollars for technology? How do we avoid medical staff conflicts as they relate to technology? How do we control equipment-related risks? and How do we maximize the useful life of the equipment or systems while minimizing the cost ownership? A hospital's clinical engineering department can assist in providing the right answers to these questions.

Technology assessment is a component of technology planning that begins with the analysis of the hospital's existing technology base. It is easy to perceive then that technology assessment, rather than an equipment comparison, is a new major function for a clinical engineering department [8]. It is important that clinical engineers be well prepared for the challenge. They must have a full understanding of the mission of their particular hospitals, a familiarity with the health care delivery system, and the cooperation of hospital administrators and the medical staff. To aid in the technology assessment process, clinical engineers need to utilize the following tools: (1) access to national database services, directories, and libraries, (2) visits to scientific and clinical exhibits, (3) a network with key industry contacts, and (4) a relationship with peers throughout the country [9].

The need for clinical engineering involvement in the technology assessment process becomes evident when recently purchased equipment or its functions are underutilized, users have ongoing problems with equipment, equipment maintenance costs become excessive, the hospital is unable to comply with standards or guidelines (i.e., JCAHO requirements) for equipment management, a high percentage of equipment is awaiting repair, or training for equipment operators is inefficient due to a shortage of allied health professionals. A deeper look at the symptoms behind these problems would likely reveal a lack of a central clearinghouse to collect, index, and monitor all technology-related information for future planning purposes, the absence of procedures for identifying emerging technologies for potential acquisition, the lack of a systematic plan for conducting technology assessment, resulting in an ability to maximize the benefits from deployment of available technology, the inability to benefit from the organization's own previous experience with a particular type of technology, the random replacement of medical technologies rather than a systematic plan based on a set of well-developed

criteria, and/or the lack of integration of technology acquisition into the strategic and capital planning of the hospital.

To address these issues, efforts to develop a technology microassessment process were initiated at one leading private hospital with the following objectives: (1) accumulate information on medical equipment, (2) facilitate systematic planning, (3) create an administrative structure supporting the assessment process and its methodology, (4) monitor the replacement of outdated technology, and (5) improve the capital budget process by focusing on long-term needs relative to the acquisition of medical equipment [10].

The process, in general, and the collection of up-to-date pertinent information, in particular, require the expenditure of certain resources and the active participation of designated hospital staff in networks providing technology assessment information. For example, corporate membership in organizations and societies that provide such information needs to be considered, as well as subscriptions to certain computerized database and printed sources [11].

At the example hospital, an MTAC was formed to conduct technology assessment. It was chaired by the director of clinical engineering. Other managers from equipment user departments usually serve as the MTAC's designated technical coordinators for specific task forces. Once the committee accepted a request from an individual user, it identified other users who might have an interest in that equipment or system and authorized the technical coordinator to assemble a task force consisting of users identified by the MTAC. This task force then took responsibility for the establishment of performance criteria that would be used during this particular assessment. The task force also should answer the questions of effectiveness, safety, and cost effectiveness as they relate to the particular assessment. During any specific period, there may be multiple task forces, each focusing on a specific equipment investigation.

The task force technical coordinator cooperates with the material management department in conducting a market survey, in obtaining the specified equipment for evaluation purposes, and in scheduling vendor-provided in-service training. The coordinator also confers with clinical staff to determine if they have experience with the equipment and the maturity level of the equipment under assessment. After establishment of a task force, the MTAC's technical coordinator is responsible for analyzing the clinical experiences associated with the use of this equipment, for setting evaluation objectives, and for devising appropriate technical tests in accord with recommendations from the task force. Only equipment that successfully passes the technical tests will proceed to a clinical trial. During the clinical trial, a task force-appointed clinical coordinator collects and reports a summary of experiences gained. The technical coordinator then combines the results from both the technical tests and the clinical trial into a summary report for MTAC review and approval. In this role, the clinical engineer/technical coordinator serves as a multidisciplinary professional, bridging the gap between the clinical and technical needs of the hospital. To complete the process, financial staff representatives review the protocol.

The technology assessment process at this example hospital begins with a department or individual filling out two forms: (1) a request for review (RR) form and (2) a capital asset request (CAR) form. These forms are submitted to the hospital's product standards committee, which determines if an assessment process is to be initiated, and the priority

for its completion. It also determines if a previously established standard for this equipment already exists (if the hospital is already using such a technology)—if so, an assessment is not needed.

On the RR, the originator delineates the rationale for acquiring the medical device. For example, the originator must tell how the item will improve quality of patient care, who will be its primary user, and how it will improve ease of use. On the CAR, the originator describes the item, estimates its cost, and provides purchase justification. The CAR is then routed to the capital budget office for review. During this process, the optimal financing method for acquisition is determined. If funding is secured, the CAR is routed to the material management department, where, together with the RR, it will be processed. The rationale for having the RR accompany the CAR is to ensure that financial information is included as part of the assessment process. The CAR is the tool by which the purchasing department initiates a market survey and later sends product requests for bid. Any request for evaluation that is received without a CAR or any CAR involving medical equipment that is received without a request for evaluation is returned to the originator without action. Both forms are then sent to the clinical engineering department, where a designated technical coordinator will analyze the requested technology maturity level and results of clinical experience with its use, review trends, and prioritize various manufactures' presentations for MTAC review.

Both forms must be sent to the MTAC if the item requested is not currently used by the hospital or if it does not conform to previously adopted hospital standards. The MTAC has the authority to recommend either acceptance or rejection of any request for review, based on a consensus of its members. A task force consisting of potential equipment users will determine the "must-have" equipment functions, review the impact of the various equipment configurations, and plan technical and clinical evaluations.

If the request is approved by the MTAC, the requested technology or equipment will be evaluated using technical and performance standards. Upon completion of the review, a recommendation is returned to the hospital's products standard committee, which reviews the results of the technology assessment, determines whether the particular product is suitable as a hospital standard, and decides if its should be purchased. If approved, the request to purchase will be reviewed by the capital budget committee (CBC) to determine if the required expenditure meets with available financial resources and if or when it may be feasible to make the purchase. To ensure coordination of the technology assessment program, the chairman of the MTAC also serves as a permanent member of the hospital's CBC. In this way, there is a planned integration between technology assessment and budget decisions.

2.4 Equipment Assets Management

An accountable, systemic approach will ensure that cost-effective, efficacious, safe, and appropriate equipment is available to meet the demands of quality patient care. Such an approach requires that existing medical equipment resources be managed and that the resulting management strategies have measurable outputs that are monitored and evaluated. Technology managers/clinical engineers are well positioned to organize and

lead this function. It is assumed that cost accounting is managed and monitored by the health care organization's financial group.

Equipment Management Process

Through traditional assets management strategies, medical equipment can be comprehensively managed by clinical engineering personnel. First, the management should consider a full range of strategies for equipment technical support. Plans may include use of a combination of equipment service providers such as manufacturers, third-party service groups, shared services, and hospital-based (in-house) engineers and biomedical equipment technicians (BMETs). All these service providers should be under the general responsibility of the technology manager to ensure optimal equipment performance through comprehensive and ongoing best-value equipment service. After obtaining a complete hospital medical equipment inventory (noting both original manufacturer and typical service provider), the management should conduct a thorough analysis of hospital accounts payable records for at least the past 2 years, compiling all service reports and preventive maintenance-related costs from all possible sources. The manager then should document in-house and external provider equipment service costs, extent of maintenance coverage for each inventory time, equipment-user operating schedule, quality of maintenance coverage for each item, appropriateness of the service provider, and reasonable maintenance costs. Next, he or she should establish an effective equipment technical support process. With an accurate inventory and best-value service providers identified, service agreements/ contracts should be negotiated with external providers using prepared terms and conditions, including a log-in system. There should be an in-house clinical engineering staff ensuring ongoing external provider cost control utilizing several tools. By asking the right technical questions and establishing friendly relationships with staff, the manager will be able to handle service purchase orders (POs) by determining if equipment is worth repairing and obtaining exchange prices for parts. The staff should handle service reports to review them for accuracy and proper use of the log-in system. They also should match invoices with the service reports to verify opportunities and review service histories to look for symptoms such as need for user training, repeated problems, run-on calls billed months apart, or evidence of defective or worn-out equipment. The manager should take responsibility for emergency equipment rentals. Finally, the manager should develop, implement, and monitor all the service performance criteria.

To optimize technology management programs, clinical engineers should be willing to assume responsibilities for technology planning and management in all related areas. They should develop policies and procedures for their hospital's management program. With life-cycle costs determined for key high-risk or high-cost devices, they should evaluate methods to provide additional cost savings in equipment operation and maintenance. They should be involved with computer networking systems within the hospital. As computer technology applications increase, the requirements to review technology-related information in a number of hospital locations will increase. They should determine what environmental conditions and facility changes are required to accommodate new technologies or changes in standards and guidelines. Lastly, they should use documentation of equipment performance and maintenance costs along with

their knowledge of current clinical practices to assist other hospital personnel in determining the best time and process for planning equipment replacement [12].

Technology Management Activities

A clinical engineering department, through outstanding performance in traditional equipment management, will win its hospital's support and will be asked to be involved in a full range of technology management activities. The department should start an equipment control program that encompasses routine performance testing, inspection, periodic and preventive maintenance, on-demand repair services, incidents investigation, and actions on recalls and hazards. The department should have multidisciplinary involvement in equipment acquisition and replacement decisions, development of new services, and planning of new construction and major renovations, including intensive participation by clinical engineering, materials management, and finance. The department also should initiate programs for training all users of patient care equipment, quality improvement (QI), as it relates to technology use, and technology-related risk management [13].

Case Study: A Focus on Medical Imaging

In the mid-1980s, a large private multihospital system contemplated the startup of a corporate clinical engineering program. The directors recognized that involvement in a diagnostic imaging equipment service would be key to the economic success of the program. They further recognized that maintenance cost reductions would have to be balanced with achieving equal or increased quality of care in the utilization of that equipment.

Programs startup was in the summer of 1987 in 3 hospitals that were geographically close. Within the first year, clinical engineering operations began in 11 hospitals in 3 regions over a two-state area. By the fall of 1990, the program included 7 regions and 21 hospitals in a five-state area. The regions were organized, typically, into teams including a regional manager and 10 service providers, serving 3 to 4 hospitals, whose average size was 225 beds. Although the staffs were stationed at the hospitals, some specialists traveled between sites in the region to provide equipment service. Service providers included individuals specializing in the areas of diagnostic imaging [x-ray and computed tomography (CT)], clinical laboratory, general biomedical instrumentation, and respiratory therapy.

At the end of the first 18 months, the program documented over $1 million in savings for the initial 11 hospitals, a 23% reduction from the previous annual service costs. Over 63% of these savings were attributable to "in-house" service x-ray and CT scanner equipment. The mix of equipment maintained by 11 imagining service providers—from a total staff of 30—included approximately 75% of the radiology systems of any kind found in the hospitals and 5 models of CT scanners from the three different manufacturers.

At the end of 3 years in 1990, program-wide savings had exceeded 30% of previous costs for participating hospitals. Within the imaging areas of the hospitals, savings approached and sometimes exceed 50% of initial service costs. The 30 imaging service

providers—out of a total staff of 62-had increased their coverage of radiology equipment to over 95%, had increased involvement with CT to include nine models from five different manufacturers, and had begun in-house work in other key imaging modalities.

Tracking the financial performance of the initial 11 hospitals over the first 3 years of the program yields of the following composite example: A hospital of 225 beds was found to have equipment service costs of $540,000 prior to program startup. Sixty-three percent of these initial costs (or $340,000) was for the maintenance of the hospital's x-ray and CT scanner systems. Three years later, annual service costs for this equipment were cut in half, to approximately $170,000. That represents a 31% reduction in hospital-wide costs due to the imaging service alone.

This corporate clinical engineering operation is, in effect, a large in-house program serving many hospitals that all have common ownership. The multihospital corporation has significant purchasing power in the medical device marketplace and provides central oversight of the larger capital expenditures for its hospitals. The combination of the parent organization's leverage and the program's commitment to serve only hospitals in the corporation facilitated the development of positive relationships with medical device manufacturers. Most of the manufacturers did not see the program as competition but rather as a potentially helpful ally in the future marketing and sales of their equipment and systems. What staff provided these results? All service providers were either medical imaging industry or military trained. All were experienced at troubleshooting electronic subsystems to component level, as necessary. Typically, these individuals had prior experience on the manufacturer's models of equipment under their coverage. Most regional managers had prior industry, third-party, or in-house imaging service management experience. Each service provider had the test equipment necessary for day-to-day duties. Each individual could expect at least 2 weeks of annual service training to keep appropriate skills current. Desired service training could be acquired in a timely manner from manufactures and/or third-party organizations. Spare or replacement parts inventory was minimal because of the program's ability to get parts from manufacturers and other sources either locally or shipped in overnight.

As quality indicators for the program, the management measured user satisfaction, equipment downtime, documentation of technical staff service training, types of user equipment errors and their effect on patient outcomes, and regular attention to hospital technology problems. User satisfaction surveys indicated a high degree of confidence in the program service providers by imaging department mangers. Problems relating to technical, management, communication, and financial issues did occur regularly, but the regional manager ensured that they were resolved in a timely manner. Faster response to daily imaging equipment problems, typically by on-site service providers, coupled with regular preventive maintenance (PM) according to established procedures led to reduced equipment downtime. PM and repair service histories were captured in a computer documentation system that also tracked service times, costs, and user errors and their effects. Assisting the safety committee became easier with ability to draw a wide variety of information quickly from the program's documenting system.

Early success in imaging equipment led to the opportunity to do some additional value-added projects such as the moving and reinstallation of x-ray rooms that preserved existing assets and opened up valuable space for installation of newer equipment and upgrades of CT scanner systems. The parent organization came to realize that these

technology management activities could potentially have a greater financial and quality impact on the hospital's health care delivery than equipment management. In the example of one CT upgrade (which was completed over two weekends with no downtime), there was a positive financial impact in excess of $600,000 and improved quality of care by allowing faster off-line diagnosis of patient scans. However, opportunity for this kind of contribution would never have occurred without the strong base of a successful equipment management program staffed with qualified individuals who receive ongoing training.

2.5 Equipment Acquisition and Deployment

2.5.1 Process of Acquiring Technology

Typically, medical device systems will emerge from the strategic technology planning and technology assessment processes as required and budgeted needs. At acquisition time, a needs analysis should be conducted, reaffirming clinical needs and device intended applications. The "request-for-review" documentation from the assessment process or capital budget request and incremental financial analysis from the planning process may provide appropriate justification information, and a capital asset request (CAR) form should be completed [14]. Materials management and clinical engineering personnel should ensure that this item is a candidate for centralized and coordinated acquisition of similar equipment with other hospital departments. Typical hospital prepurchase evaluation guidelines include an analysis of needs and development of a specification list, formation of a vendor list and requesting proposals, analyzing proposals and site planning, evaluating samples, selecting finalists, making the award, delivery and installation, and acceptance testing. Formal request for proposals (RFPs) from potential equipment vendors are required for intended acquisitions whose initial or life-cycle cost exceeds a certain threshold, i.e., $100,000. Finally, the purchase takes place, wherein final equipment negotiations are conducted and purchase documents are prepared, including a purchase order.

Acquisition Process Strategies

The cost-of-ownership concept can be used when considering what factors to include in cost comparisons of competing medical devices. Cost of ownership encompasses all the direct and indirect expenses associated with medical equipment over its lifetime [15]. It expresses the cost factors of medical equipment for both the initial price of the equipment (which typically includes the equipment, its installation, and initial training cost) and over the long term. Long-term costs include ongoing training, equipment service, supplies, connectivity, upgrades, and other costs. Health care organizations are just beginning to account for a full range of cost-of-ownership factors in their technology assessment and acquisition processes, such as acquisition costs, operating costs, and maintenance costs (installation, supplies, downtime, training, spare parts, test equipment and tools, and depreciation). It is estimated that the purchase price represents only 20% of the life-cycle cost of ownership.

When conducting needs analysis, actual utilization information from the organization's existing same or similar devices can be very helpful. One leading private multihospital system has implemented the following approach to measuring and developing relevant management feedback concerning equipment utilization. It is conducting equipment utilization review for replacement planning, for ongoing accountability of equipment use, and to provide input before more equipment is purchased. This private system attempts to match product to its intended function and to measure daily (if necessary) the equipment's actual utilization. The tools they use include knowing their hospital's entire installed base of certain kinds of equipment, i.e., imaging systems. Utilization assumptions for each hospital and its clinical procedural mix are made. Equipment functional requirements to meet the demands of the clinical procedures are also taken into account.

Life-cycle cost analysis is a tool used during technology planning, assessment, or acquisition "either to compare high-cost, alternative means for providing a service or to determine whether a single project or technology has a positive or negative economic value. The strength of the life-cycle cost analysis is that it examines the cash flow impact of an alternative over its entire life, instead of focusing solely on initial capital investments" [15].

"Life-cycle cost analysis facilitates comparisons between projects or technologies with large initial cash outlays and those with level outlays and inflows over time. It is most applicable to complex, high-cost choices among alternative technologies, new service, and different means for providing a given service. Life-cycle cost analysis is particularly useful for decisions that are too complex and ambiguous for experience and subjective judgment alone. It also helps decision makers perceive and include costs that often are hidden or ignored, and that may otherwise invalidate results" [12].

"Perhaps the most powerful life-cycle cost technique is net present value (NPV) analysis, which explicitly accounts for inflation and foregone investment opportunities by expressing future cash flows in present dollars" [12].

Examples where LCC and NPV analysis prove very helpful are in deciding whether to replace/rebuild or buy/lease medical imaging equipment. The kinds of costs captured in life-cycle cost analysis, include decision-making costs, planning agency/certificate of need costs (if applicable), financing, initial capital investment costs including facility changes, life-cycle maintenance and repairs costs, personnel costs, and other (reimbursement consequences, resale, etc.).

One of the best strategies to ensure that a desired technology is truly of value to the hospital is to conduct a careful analysis in preparation for its assimilation into hospital operations. The process of equipment prepurchase evaluation provides information that can be used to screen unacceptable performance by either the vendor or the equipment before it becomes a hospital problem.

Once the vendor has responded to informal requests or formal RFPs, the clinical engineering department should be responsible for evaluating the technical response, while the materials management department should devaluate the financial responses.

In translating clinical needs into a specification list, key features or "must-have" attributes of the desired device are identified. In practice, clinical engineering and materials management should develop a must-have list and an "extras" list. The extras list contains features that may tip the decision in favor of one vendor, all other factors being

even. These specification lists are sent to the vendor and are effective in a self-elimination process that results in a time savings for the hospital. Once the “must-have” attributes have been satisfied, the remaining candidate devices are evaluated technically, and the extras are considered. This is accomplished by assigning a weighting factor (i.e., 0 to 5) to denote the relative importance of each of the desired attributes. The relative ability of each device to meet the defined requirements is then rated [15].

One strategy that strengthens the acquisition process is the conditions-of-sale document. This multifaceted document integrates equipment specifications, performance, installation requirements, and follow-up services. The conditions-of-sale document ensures that negotiations are completed before a purchase order is delivered and each participant is in agreement about the product to be delivered. As a document of compliance, the conditions-of-sale document specifies the codes and standards having jurisdiction over that equipment. This may include provisions for future modification of the equipment, compliance with standards under development, compliance with national codes, and provision for software upgrades.

Standard purchase orders that include the conditions of sale for medical equipment are usually used to initiate the order. At the time the order is placed, clinical engineering is notified of the order. In addition to current facility conditions, the management must address installation and approval requirements, responsibilities, and timetable; payment, assignment, and cancellation; software requirements and updates; documentation; clinical and technical training; acceptance testing (hospital facility and vendor); warranty, spare parts, and service; and price protection.

All medical equipment must be inspected and tested before it is placed into service regardless of whether it is purchased, leased, rented, or borrowed by the hospital. In any hospital, clinical engineering should receive immediate notification if a very large device or system is delivered directly into another department (e.g., imaging or cardiology) for installation. Clinical engineering should be required to sign off on all purchase orders for devices after installation and validation of satisfactory operation. Ideally, the warranty period on new equipment should not begin until installation and acceptance testing are completed. It is not uncommon for a hospital to lose several months of free parts and service by the manufacturer when new equipment is, for some reason, not installed immediately after delivery.

Clinical Team Requirements

During the technology assessment and acquisition processes, clinical decision-makers analyze the following criteria concerning proposed technology acquisitions, specifically as they relate to clinical team requirements: ability of staff to assimilate the technology, medical staff satisfaction (short term and long term), impact on staffing (numbers, functions), projected utilization, ongoing related supplies required, effect on delivery of care and outcomes (convenience, safety, or standard of care), result of what is written in the clinical practice guidelines, credentialling of staff required, clinical staff initial and ongoing training required, and the effect on existing technology in the department or on other services/departments.

Defining Terms

Appropriate technology [1]: A term used initially in developing countries, referring to selecting medical equipment that can "appropriately" satisfy the following constraints: funding shortages, insufficient numbers of trained personnel, lack of technical support, inadequate supplies of consumables/ accessories, unreliable water an power utilities/supplies, and lack of operating and maintenance manuals. In the context of this chapter, appropriate technology selection must taken into consideration local health needs and disease prevalence, the need for local capability of equipment maintenance, and availability of resources for ongoing operational and technical support.

Clinical engineers/biomedical engineers: As we began describing the issues with the management of medical technology, it became obvious that some of the terms are being used interchangeably in the literature. For example, the terms *engineers, clinical engineers, biomedical equipment technicians, equipment managers,* and *health care engineers* are frequently used. For clarification, in this chapter we will refer to clinical engineers and the clinical engineering department as a representative group for all these terms.

Cost effectiveness [1]: A mixture of quantitative and qualitative considerations. It includes the health priorities of the country or region at the macro-assessment level and the community needs at the institution micro-assessment level. Product life-cycle cost analysis (which, in turn, includes initial purchase price, shipping, renovations, installation, supplies, associated disposables, cost per use, and similar quantitative measures) is a critical analysis measure. Life-cycle cost also takes into account staff training, ease of use, service, and many other cost factors. But experience and judgment about the relative importance of features and the ability to fulfill the intended purpose also contribute critical information to the cost-effectiveness equation.

Equipment acquisition and deployment: Medical device systems and products typically emerge from the strategic technology planning process as "required and budgeted" needs. The process that follows, which ends with equipment acceptance testing and placement into general use, is known as the *equipment acquisition and deployment process*.

Health care technology: Health care technology includes the devices, equipment, systems, software, supplies, pharmaceuticals, biotechnologies, and medical and surgical procedures used in the prevention, diagnosis, and treatment of disease in humans, for their rehabilitation, and for assistive purposes. In short, technology is broadly defined as encompassing virtually all the human interventions intended to cope with disease and disabilities, short of spiritual alternatives. This chapter focuses on medical equipment products (devices, systems, and software) rather than pharmaceuticals, biotechnologies, or procedures [1]. The concept of technology also encompasses the facilities that house both patients and products. Facilities cover a wide spectrum—from the modern hospital on one end to the mobile imaging trailer on the other.

Quality of care (QA) and quality of improvement (QI): Quality assurance (QA) and Quality improvement (QI) are formal sets of activities to measure the quality of care provided; these usually include a process for selecting, monitoring, and applying corrective measures. The 1994 Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) standards require hospital QA, programs to focus on patient outcomes as a primary reference. JCAHO standards for plant, technology, and safety

management (PTSM), in turn, require certain equipment management practices and QA or QI activities. Identified QI deficiencies may influence equipment planning, and QI audits may increase awareness of technology overuse or under utilization.

Risk management: Risk management is a program that helps the hospital avoid the possibility of risks, minimize liability exposure, and stay compliant with regulatory reporting requirements. JCAHO PTSM standards require minimum technology-based risk-management activities. These include clinical engineering's determination of technology-related incidents with follow-up steps to prevent recurrences and evaluation and documentation of the effectiveness of these steps.

Safety: Safety is the condition of being safe from danger, injury, or damage. It is judgment about the acceptability of risk in a specified situation (e.g., for a given medical problem) by a provider with specified training at a specified type of facility equipment.

Standards [1]: A wide variety of formal standards and guidelines related to health care technology now exists. Some standards apply to design, development, and manufacturing practices for devices, software, and pharmaceuticals; some are related to the construction and operation of a health care facility; some are safety and performance requirements for certain classes of technologies, such as standards related to radiation or electrical safety; and others relate to performance, or even construction specifications, for specific types of technologies. Other standards and guidelines deal with administrative, medical, and surgical procedures and the training of clinical personnel. Standards and guidelines are produced and/or adopted by government agencies, international organizations, and professional and specialty organizations and societies. ECRI's *Healthcare Standards Directory* lists over 20,000 individual standards and guidelines produced by over 600 organizations and agencies from North America alone.

Strategic technology planning: Strategic technology planning encompasses both technologies new to the hospital and replacements for existing equipment that are to be acquired over several quarters. Acquisitions can be proposed for reasons related to safety, standard-of-care issues, and age or obsolescence of existing equipment. Acquisitions also can be proposed to consolidate several service areas, expand a service areas to reduce cost of service, or add a new service area.

Strategic technology panning optimizes the way the hospital's capital resources contribute to its mission. It encourages choosing new technologies that are cost-effective, and it also allows the hospital to be competitive in offering state-of-the-art services. Strategic technology planning works for a single department, product line, or clinical service. It can be limited to one or several high-priority areas. It also can be used for an entire mulihospital system or geographic region [4].

Technology assessment: Assessment of medical technology is any process used for examining and reporting properties of medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended [2]. A primary technology assessment is one that seeks new, previously nonexistent data through research, typically employing long-term clinical studies of the type described below. A secondary technology assessment is usually based on published data, interviews, questionnaires, and other information-gathering methods rather than original research that creates new, basic data.

In technology assessment, there are six basic objectives that the clinical engineering department should have in mind. First, there should be ongoing monitoring of developments concerning new and emerging technologies. For new technologies, there should be an assessment of the clinical efficacy, safety, and cost/benefit ratio, including their effects on established technologies. There should be an evaluation of the short- and long-term costs and benefits of alternate approaches to managing specific clinical conditions. The appropriateness of existing technologies and their clinical uses should be estimated, while outmoded technologies should be identified and eliminated from their duplicative uses. The department should rate specific technology-based interventions in terms of improved overall value (quality and outcomes) to patients, providers, and payers. Finally, the department should facilitate a continuous uniformity between needs, offerings, and capabilities [3],

The locally based (hospital or hospital group) technology assessment described in this chapter is a process of secondary assessment that attempts to judge whether a certain medical equipment/ product can be assimilated into the local operational environment.

Technology diffusion [1]: The process by which a technology is spread over time in a social system. The progression of technology diffusion can be described in four stages. The *emerging* or applied research stage occurs around the time of initial clinical testing. In the *new* stage, the technology has passed the phase of clinical trials but is not yet in widespread use. During the *established* stage, the technology is considered by providers to be a standard approach to a particular condition and diffuses into general use. Finally, in the *obsolete/outmoded* stage, the technology is superseded by another and/or is demonstrated to be ineffective or harmful.

Technology life cycle: Technology has a life cycle—a process by which technology is crated, tested, applied, and replaced or abandoned. Since the life cycle varies from basic research and innovation to obsolescence and abatement, it is critical to know the maturity of a technology prior to making decisions regarding its adoption. Technology forecast assessment of pending technological changes are the investigative tools that support systematic and rational decisions about the utilization of a given institution's technological capabilities.

Technology planning and management [3]: Technology planning and management are an accountable, systematic approach to ensuring that cost-effective, efficacious, appropriate, and safe equipment is available to meet the demands of quality patient care and allow an institution to remain competitive. Elements include in-house service management, management and analysis of equipment external service providers, involvement in the equipment acquisition process, involvement of appropriate hospital personnel in facility planning and design, involvement in reducing technol-ogy-related patient and staff incidents, training equipment users, reviewing equipment replacement needs, and ongoing assessment of emerging technologies [4].

References

1. ECRI. *Healthcare Technology Assessment Curriculum.* Philadelphia, August 1992.
2. Banata HD, Institute of Medicine. *Assessing Medical Technologies.* Washington, National Academy Press, 1985.

3. Lumsdon K. Beyond technology assessment: Balancing strategy needs, strategy. *Hospitals* 15:25, 1992.
4. ECRI. *Capital, Competition, and Constraints: Managing Healthcare in the 1990s.* A Guide for Hospital Executives. Philadelphia, 1992.
5. Berkowtiz DA, Solomon RP. Providers may be missing opportunities to improve patient outcomes. *Costs, Outcomes Measurement and Management* May-June: 7, 1991.
6. ECRI. *Regional Healthcare Technology Planning and Management Program.* Philadelphia, 1990.
7. Sprague GR. Managing technology assessment and acquisition. *Health Exec* 6:26, 1988.
8. David Y. Technology-related decision-making issues in hospitals. In *IEEE Engineering in Medicine and Biology Society Proceedings of the llth Annual International Conference,* 1989.
9. Wagner M. Promoting hospitals high-tech equipment. *Mod Healthcare* 46, 1989.
10. David Y. *Medical Technology 2001.* CPA Healthcare Conference, 1992.
11. ECRI. *Special Report on Technology Management,* Health Technology. Philadelphia, 1989.
12. ECRI. *Special Report on Devices and Dollars,* Philadelphia, 1988.
13. Gullikson ML, David Y, Brady MH. An automated risk management tool. *JCAHO, Plant, Technology and Safety Management Review,* PTSM Series, no 2, 1993.
14. David Y, Judd T, ECRI. *Special Report on Devices and Dollars,* Philadelphia, 1988. *Medical Technology Management.* SpaceLabs Medical, Inc., Redmond, Wash 1993.
15. Bronzino JD (ed). *Management of Medical Technology: A Primer for Clinical Engineers.* Stoneham, Mass, Butterworth, 1992.
16. David Y. *Risk Measurement For Managing Medical Technology.* Conference Proceedings, PERM-IT 1997, Australia.

3
Risk Factors, Safety, and Management of Medical Equipment

Michael L.Gullikson
Texas Children's Hospital

3.1 Risk Management: A Definition

Inherent in the definition of *risk management* is the implication that the hospital environment cannot be made risk-free. In fact, the nature of medical equipment—to invasively or noninvasively perform diagnostic, therapeutic, corrective, or monitoring intervention on behalf of the patient—implies that risk is present. Therefore, a standard of acceptable risk must be established that defines *manageable risk* in a real-time economic environment.

Unfortunately, a preexistent, quantitative standard does not exist in terms of, for instance, mean time before failure (MTBF), number of repairs or repair redos per equipment item, or cost of maintenance that provides a universal yardstick for risk management of medical equipment. Sufficient clinical management of risk must be in place that can utilize safeguards, preventive maintenance, and failure analysis information to minimize the occurrence of injury or death to patient or employee or property damage. Therefore, a process must be put in place that will permit analysis of information and modification of the preceding factors to continuously move the medical equipment program to a more stable level of manageable risk.

Risk factors that require management can be illustrated by the example of the "double-edged" sword concept of technology (see Fig. 3.1). The front edge of the sword represents the cutting edge of technology and its beneficial characteristics: increased quality, greater availability of technology, timeliness of test results and treatment, and so on. The back edge of the sword represents those liabilities that must be addressed to effectively manage risk: the hidden costs discussed in the next paragraph, our dependence on technology, incompatibility of equipment, and so on [1].

For example, the purchase and installation of a major medical equipment item may only represent 20% of the lifetime cost of the equipment [2]. If the operational budget of a nursing floor does not include the other 80% of the equipment costs, the budget constraints may require cutbacks where they appear to minimally affect direct patient care. Preventive maintenance, software upgrades that address "glitches," or overhaul requirements, may be seen as unaffordable luxuries. Gradual equipment deterioration without maintenance may bring the safety level below an acceptable level of manageable risk.

0-8493-1813-0/03/$0.00+$1.50

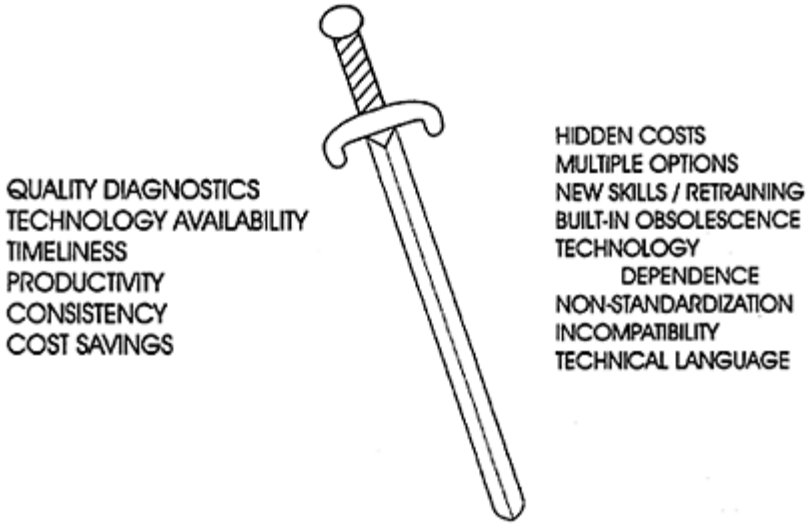


FIGURE 3.1 Double-edged sword concept of risk management.

Since economic factors as well as those of safety must be considered, a balanced approach to risk management that incorporates all aspects of the medical equipment lifecycle must be considered.

The operational flowchart in Fig. 3.2 describe the concept of medical equipment life-cycle management from the clinical engineering department viewpoint. The flowchart includes planning, evaluation, and initial purchase documentation requirements. The condition of sale, for example, ensures that technical manuals, training, replacement parts, etc. are received so that all medical equipment might be fully supported in-house after the warranty period. Introduction to the preventive maintenance program, unscheduled maintenance procedures, and retirement justification must be part of the process. Institutional-wide cooperation with the life-cycle concept requires education and patience to convince health care providers of the team approach to managing medical equipment technology.

This balanced approach requires communication and comprehensive planning by a health care team responsible for evaluation of new and shared technology within the organization. A medical technology evaluation committee (see Fig. 3.3), composed of representatives from administration, medical staff, nursing, safety department, biomedical engineering, and various services, can be an effective platform for the integration of technology and health care. Risk containment is practiced as the committee reviews not only the benefits of new technology but also the technical and clinical liabilities and provides a 6-month follow-up study to measure the effectiveness of the selection process. The history of risk management in medical equipment management provides helpful insight into its current status and future direction.

3.2 Risk Management: Historical Perspective

Historically, risk management of medical equipment was the responsibility of the clinical engineer (Fig. 3.4). The engineer selected medical equipment based on individual clinical department consultations and established preventive maintenance (PM) programs based on manufacturer's recommendation and clinical experience. The clinical engineer reviewed the documentation and "spot-checked" equipment used in the hospital. The clinical engineer met with biomedical supervisors and technicians to discuss PM completion and to resolve repair problems. The clinical engineer then attempted to analyze failure information to avoid repeat failure.

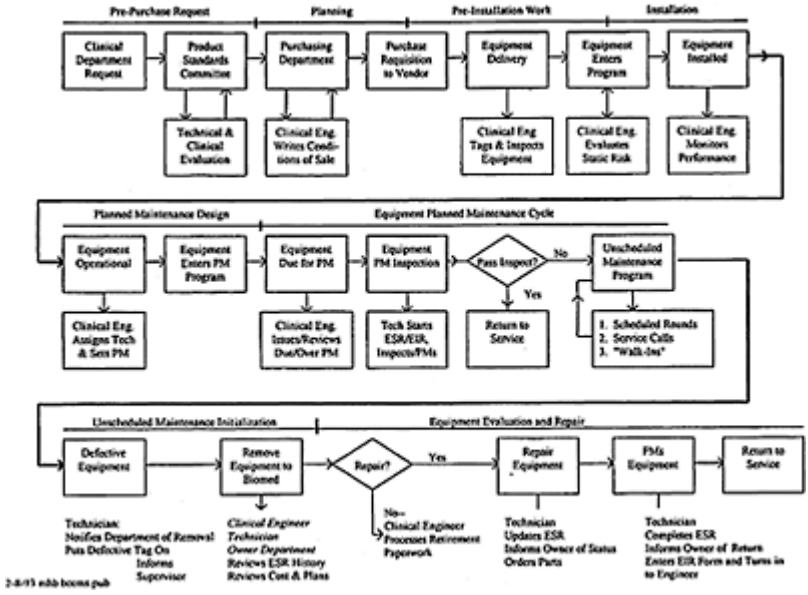


FIGURE 3.2 Biomedical engineering equipment management system (BEEMS).

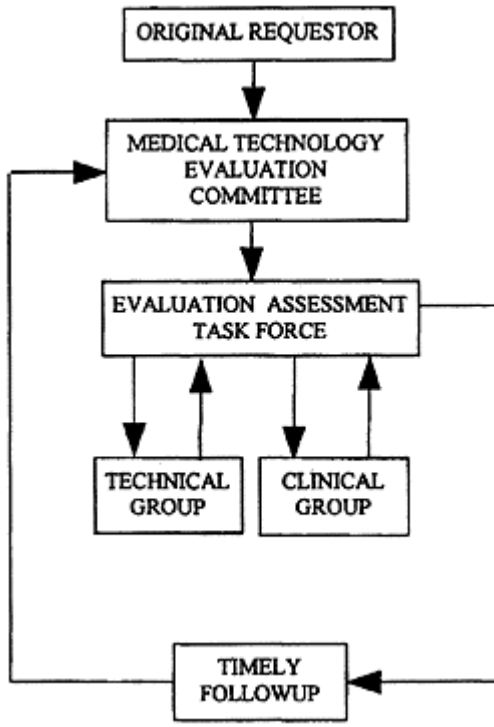


FIGURE 3.3 Medical technology evaluation committee.

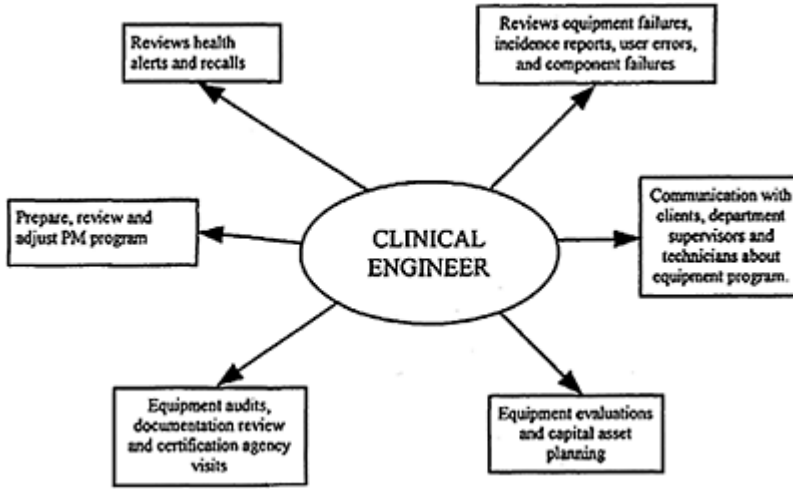


FIGURE 3.4 Operational flowchart.

However, greater public awareness of safety issues, increasing equipment density at the bedside, more sophisticated software-driven medical equipment, and financial considerations have made it more difficult for the clinical engineer to singularly handle risk issues. In addition, the synergistic interactions of various medical systems operating in proximity to one another have added another dimension to the risk formula. It is not only necessary for health care institutions to manage risk using a team approach, but it is also becoming apparent that the clinical engineer requires more technology-intensive tools to effectively contribute to the team effort [3].

100 Medical Equipment Operator Error

101 Medical Equipment Failure

102 Medical Equipment Physical Damage

103 Reported Patient Injury

104 Reported Employee Injury

105 Medical Equipment Failed PM

108 Medical Equipment MBA

FIGURE 3.5 Failure codes.

3.3 Risk Management: Strategies

Reactive risk management is an outgrowth of the historical attitude in medical equipment management that risk is an anomaly that surfaces in the form of a failure. If the failure is analyzed and proper operational procedures, user in-services, and increased maintenance are supplied, the problem will disappear and personnel can return to their normal work. When the next failure occurs, the algorithm is repeated. If the same equipment fails, the algorithm is applied more intensely. This is a useful but not comprehensive component of risk management in the hospital. In fact, the traditional methods of predicting the reliability of electronic equipment from field failure data have not been very effective [4]. The health care environment, as previously mentioned, inherently contains risk that must be maintained at a manageable level. A reactive tool cannot provide direction to a risk-management program, but it can provide feedback as to its efficiency.

The engine of the reactive risk-management tool is a set of failure codes (see Fig. 3.5) that flag certain anomalous conditions in the medical equipment management program. If operator training needs are able to be identified, then codes 100, 102, 104, and 108 (MBA equipment returned within 9 days for a subsequent repair) may be useful. If technician difficulties in handling equipment problems are of concern, then 108 may be of interest. The key is to develop failure codes, not in an attempt to define all possible anomaly modalities, but for those that *can clearly be defined and provide unambiguous direction for the correction process*. Also, the failure codes should be linked to equipment type, manufacturer/model, technician service group, hospital, and clinical department. Again, when the data are analyzed, will the result be provided to an administrator, engineer, clinical departmental director, or safety department? This should determine the format in which the failure codes are presented.

A report intended for the clinical engineer might be formatted as in Fig. 3.6. It would consist of two parts, sorted by equipment type and clinical department (not shown). The engineer's report shows the failure code activity for various types of equipment and the distribution of those failure codes in clinical departments.

Additionally, fast data-analysis techniques introduced by NASA permit the survey of large quantities of information in a three-dimensional display [5] (Fig. 3.7). This approach permits viewing time-variable changes from month to month and failure concentration in specific departments and equipment types.

The importance of the format for failure modality presentation is critical to its usefulness and acceptance by health care professionals. For instance, a safety director requests the clinical engineer to provide a list of equipment that, having failed, could have potentially harmed a patient or employee. The safety director is asking the clinical engineer for a clinical judgment based on clinical as well as technical factors. This is beyond the scope of responsibility and expertise of the clinical engineer. However, the request can be addressed indirectly. The safety director's request can be addressed in two steps: first, providing a list of high-risk equipment (assessed when the medical equipment is entered into the equipment inventory) and, second, a clinical judgment based on equipment failure mode, patient condition, and so on. The flowchart in Fig. 3.8 provides the safety director with useful information but does not require the clinical engineer to make an unqualified clinical judgment. If the "failed PM" failure code were selected

Source	Total PM	Fail PM Items	% Fail	Reported Fail-OK	Physical Damage	Patient Injury	Employee Injury	Back Again	Equip Fail	Equip Count	% Fail
10514 PULMONARY INSTR TEXAS CHILDREN'S HOSP											
1 NON-TAGGED EQUIPMENT	0		0.00	1	5				14		0.00
1200 THERMOMETER, ELECTRONIC	2		0.00						1	99	1.01
1292 RADIANT WARMER, INFANT	1		0.00						3	43	6.76
1306 INCUBATOR, NEONATAL	0	2	0.00		7				9	56	16.07
1307 INCUBATOR, TRANSPORT, NEONATAL	9	3	33.33		1			1 *	4	9	44.44
1320 PHOTOTHERAPY UNIT, NEONATAL	4		0.00						2	28	7.14
1321 INFUSION PUMP	35		0.00	15	9			3 *	34	514	6.61
1357 SUCTION-UNIT POWERED,BODY FLUID	0		0.00						4	358	1.12
1384 EXAMINATION LIGHT, AC-POWERED	11		0.00						1	47	2.13
1447 CARDIAC MONITOR W/ RATE ALARM	0		0.00	1	1			1 *	1		0.00
1567 SURGICAL NERVE STIMULATOR /LOC	0		0.00						2	15	13.33
1624 STOSCOPE	73		0.00						1	101	0.99
1675 OXYGEN GAS ANALYZER	0		0.00						3	46	6.52
1681 SPIROMETER DIAGNOSTIC	0		0.00						1	8	12.50
1703 AIRWAY PRESSURE MONITOR	1		0.00					1 *	7	9	77.78
1735 BREATHING GAS MIXER	13		0.00					1 *	4	28	15.63
1749 HYPO/HYPERTHERMIA DEVICE	1		0.00						1	3	33.33
1762 NEBULIZER	25		0.00						1	54	1.79
1787 VENTILATOR CONTINUOUS	96	7	7.29		1			1 *	11	99	11.11
1788 VENTILATOR MOMENTINUOUS	1		0.00		1				3	27	11.11
2014 HEMODIALYSIS SYSTEM ACCESSORIE	0		0.00						3	1	300.00
2051 PERITONEAL DIALYSIS SYS & ACC	4		0.00						1	5	20.00
2084 SPECTROPHOTOMETER, MASS	0		0.00						2	3	66.67
2095 POWERED SUCTION PUMP	14		0.00		1			1 *	1	32	3.13
5028 PH METER	0		0.00		1				1	4	25.00
5035 COMPUTER & PERIPHERALS	0		0.00						3	18	16.67
5081 OXYGEN MONITOR	0		0.00	3	1				11	150	14.71
5082 RESPIRATION ANALYZER	1		0.00						1	5	20.00
5097 EGWA TABLE	75		0.00						1	84	1.14
5113 PRINTER	2		0.00						1	12	8.33
5126 ARCHOSCOGRAPH	2		0.00						4	1	400.00
9108 STADIOMETER	0		0.00						1	8	12.50
17211 ANESTHESIA MONITOR	23		0.00						2	23	8.70
90063 POWER SUPPLY, PORTABLE	29		0.00						1	25	4.00
Total for TEXAS CHILDREN'S HOSP	415	12		20	28			9	144	1899	

FIGURE 3.6 Engineer's failure analysis report.

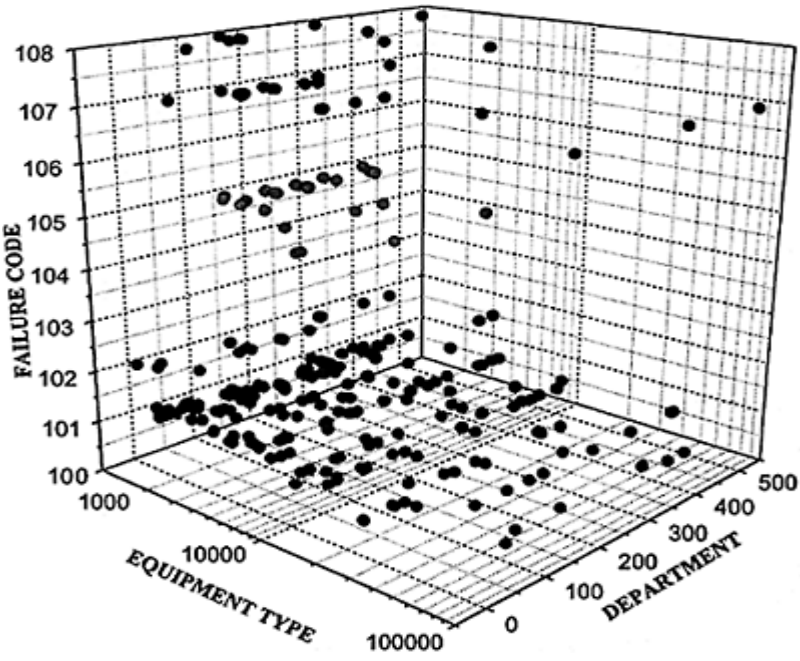


FIGURE 3.7 Failure code analysis using a 3-D display.

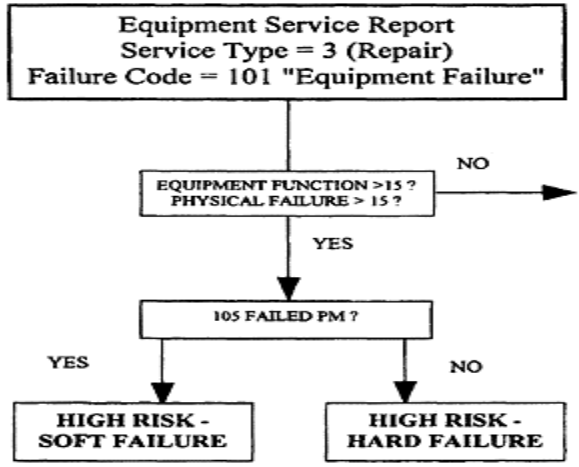


FIGURE 3.8 High-risk medical equipment failures.

from the list of high-risk medical equipment requiring repair, the failure would be identified by the technician during routine preventive maintenance, and most likely the clinician still would find the equipment clinically efficacious. This condition is a "high-risk, soft-failure" or a high-risk equipment item whose failure is *least* likely to cause injury. If the "failed PM" code were not used, the clinician would question the clinical efficacy of the medical equipment item, and the greater potential for injury would be identified by "high-risk, hard-failure." Monitoring the distribution of high-risk equipment in these two categories assists the safety director in managing risk.

Obviously, a more forward-looking tool is needed to take advantage of the failure codes and the plethora of equipment information available in a clinical engineering department. This proactive tool should use failure codes, historical information, the "expert" knowledge of the clinical engineer, and the baseline of an established "manageable-risk" environment (perhaps not optimal but stable).

The overall components and process flow for a proactive risk-management tool [6] are presented in Fig. 3.9. It consists of a two-component static risk factor, a two-component dynamic risk factor, and two "shaping" or feedback loops.

The static risk factor classifies new equipment by a generic equipment type: defibrilator, electrocardiograph, pulse oximeter, etc. When equipment is introduced into the equipment database, it is assigned to two different static risk (Fig. 3.10) categories [7]. The first is the equipment function that defines the application and environment in which the equipment item will operate. The degree of interaction with the patient is also taken into account. For example, a therapeutic device would have a higher risk assignment than a monitoring or diagnostic device. The second component of the static risk factor is the physical risk category. It defines the worst-case scenario in the event of equipment malfunction. The correlation between equipment function and physical risk on many items might make the two categories appear redundant. However, there are sufficient equipment types where this is not the case. A scale of 1 to 25 is assigned to each risk category. The larger number is assigned to devices demonstrating greater risk because of their function or the consequences of device failure. The 1-to-25 scale is an arbitrary assignment, since a validated scale of risk factors for medical equipment, as previously described, is nonexistent. The risk points assigned to the equipment from these two categories are algebraically summed and designated the static risk factor. This value remains with the equipment type and the individual items within that equipment type permanently. Only if the equipment is used in a clinically variant way or relocated to a functionally different environment would this assignment be reviewed and changed.

The dynamic component (Fig. 3.11) of the risk-management tool consists or two parts. The first is a maintenance requirement category that is divided into 25 equally spaced divisions, ranked by least (1) to greatest (25) average manhours per device per year. These divisions are scaled by the maintenance hours for the equipment type requiring the greatest amount of maintenance attention. The amount of non-planned (repair) manhours from the previous 12 months of service reports is totaled for each equipment type. Since this is maintenance work on failed equipment items, it correlates with the risk associated with that equipment type.

If the maintenance hours of an equipment type are observed to change to the point of placing it in a different maintenance category, a flag notifies the clinical engineer to review the equipment-type category. The engineer may increase the PM schedule to

compensate for the higher unplanned maintenance hours. If the engineer believes the system "overacted," a "no" decision adjusts a scaling factor by −5%. Progressively, the algorithm is "shaped" for the equipment maintenance program in that particular institution. However, to ensure that critical changes in the average manhours per device for each equipment type is not missed during the shaping period, the system is initialized. This is accomplished by increasing the average manhours per device for each equipment type to within 5% of the next higher maintenance requirement division. Thus the system is sensitized to variations in maintenance requirements.

The baseline is now established for evaluating individual device risk. Variations in the maintenance requirement hours for any particular equipment type will, for the most part, only occur over a substantial period of time. For this reason, the maintenance requirement category is designated a "slow" dynamic risk element.

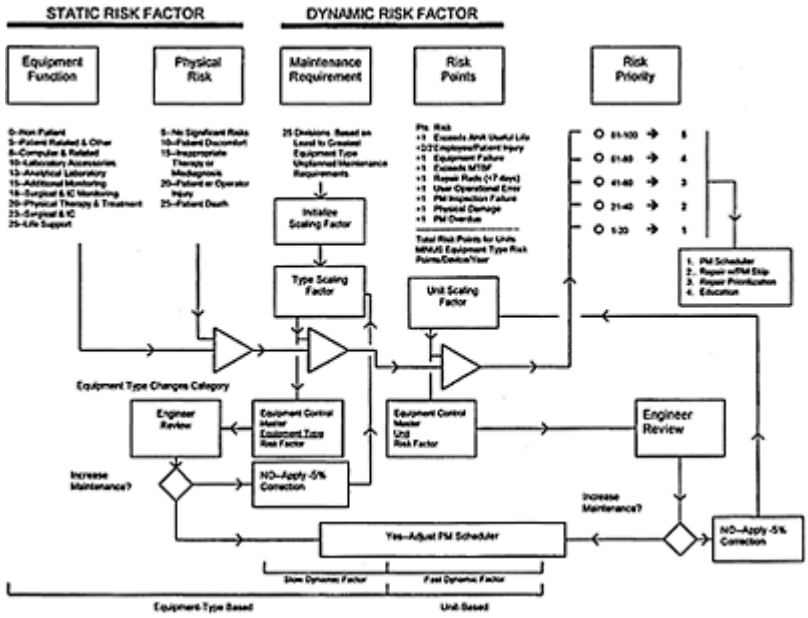


FIGURE 3.9 Biomedical engineering risk-management tool.

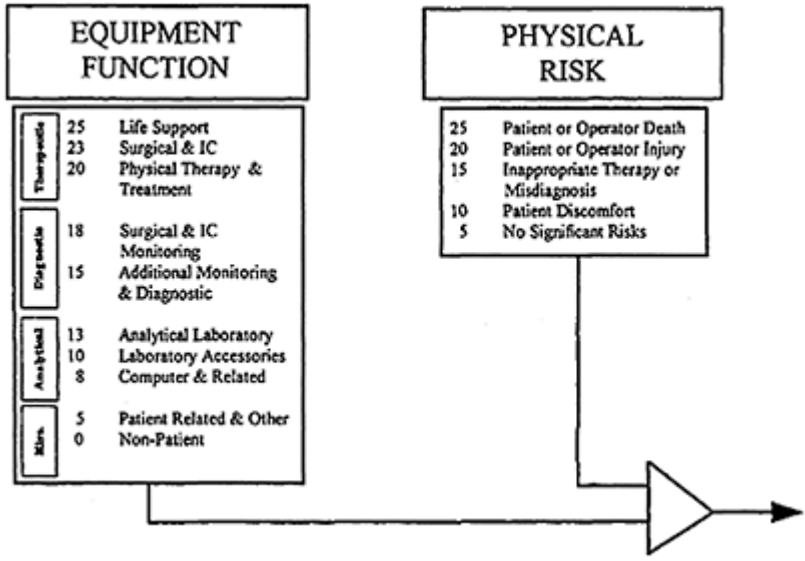


FIGURE 3.10 Static risk components.

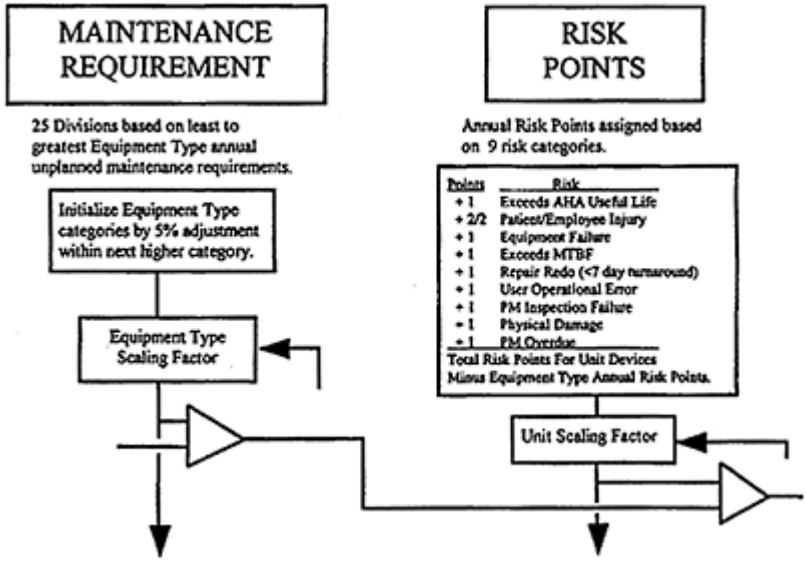


FIGURE 3.11 Dynamic risk components.

The second dynamic element assigns weighted risk points to *individual* equipment items for each unique risk occurrence. An *occurrence* is defined as when the device

- Exceeds the American Hospital Association Useful Life Table for Medical Equipment or exceeds the historical MTBF for that manufacturer and model
- Injures a patient or employee
- Functionally fails or fails to pass a PM inspection
- Is returned for repair or returned for rerepair within 9 days of a previous repair occurrence
- Misses a planned maintenance inspection
- Is subjected to physical damage
- Was reported to have failed but the problem was determined to be a user operational error

These risk occurrences include the failure codes previously described. Although many other risk occurrences could be defined, these nine occurrences have been historically effective in managing equipment risk. The risk points for each piece or equipment are algebraically summed over the previous year. Since the yearly total is a moving window, the risk points will not continue to accumulate but will reflect a recent historical average risk. The risk points for each equipment type are also calculated. This provides a baseline to measure the relative risk of devices within an equipment type. The average risk points for the equipment type are subtracted from those for each piece of equipment within the equipment type. If the device has a negative risk point value, the device's risk is less than the average device in the equipment type. If positive, then the device has higher risk than the average device. This positive or negative factor is algebraically summed to the risk values from the equipment function, physical risk, and maintenance requirements. The annual risk points for an individual piece of equipment might change quickly over several months. For this reason, this is the "fast" component of the dynamic risk factor.

The concept of risk has now been quantified in term of equipment function, physical risk, maintenance requirements, and risk occurrences. The total risk count for each device then places it in one of five risk priority groups that are based on the sum of risk points. These groups are then applied in various ways to determine repair triage, PM triage, educational and in-service requirements and test equipment/parts, etc. in the equipment management program.

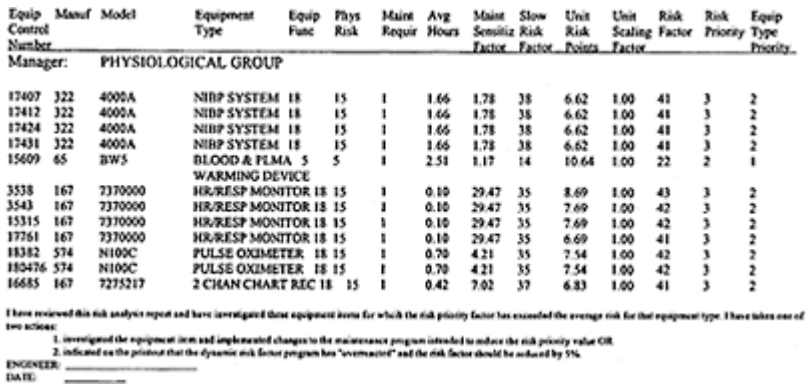
Correlation between the placement of individual devices in each risk priority group and the levels of planned maintenance previously assigned by the clinical engineer have shown that the proactive risk-management tool calculates a similar maintenance schedule as manually planned by the clinical engineer. In other words, the proactive risk-management tool algorithm places equipment items in a risk priority group commensurate with the greater or lesser maintenance as currently applied in the equipment maintenance program.

As previously mentioned, the four categories and the 1 to 25 risk levels within each category are arbitrary because a "gold standard" for risk management is nonexistent. Therefore, the clinical engineer is given input into the dynamic components making up the risk factor to "shape the system" based on the equipment's maintenance history and the clinical engineer's experience. Since the idea of a safe medical equipment program involves "judgment about the acceptability of risk in a specified situation" [8], this experience is a necessary component of the risk-assessment tool for a specific health care setting.

In the same manner, the system tracks the unit device's assigned risk priority group. If the risk points for a device change sufficiently to place it in a different group, it is flagged for review. Again, the clinical engineer reviews the particular equipment item and decides if corrective action is prudent. Otherwise, the system reduces the scaling factor by 5%. Over a period of time, the system will be "formed" to what is acceptable risk and what deserves closer scrutiny.

3.4 Risk Management: Application

The information can be made available to the clinical engineer in the form of a risk assessment report (see Fig. 3.12). The report lists individual devices by property tag number (equipment control number),



Equip Control Number	Manuf	Model	Equipment Type	Equip Func	Phys Risk	Maint Requir	Avg Hours	Maint Sensitiz Factor	Slow Risk Factor	Unit Risk Points	Unit Scaling Factor	Risk Factor	Risk Priority	Equip Type Priority
Manager:		PHYSIOLOGICAL GROUP												
17407	322	4000A	NIBP SYSTEM	18	15	1	1.66	1.78	38	6.62	1.00	41	3	2
17412	322	4000A	NIBP SYSTEM	18	15	1	1.66	1.78	38	6.62	1.00	41	3	2
17424	322	4000A	NIBP SYSTEM	18	15	1	1.66	1.78	38	6.62	1.00	41	3	2
17431	322	4000A	NIBP SYSTEM	18	15	1	1.66	1.78	38	6.62	1.00	41	3	2
15609	65	BW5	BLOOD & FLMA WARMING DEVICE	5	5	1	2.51	1.17	14	10.64	1.00	22	2	1
3538	167	7370000	HR/RESP MONITOR	18	15	1	0.10	29.47	35	8.69	1.00	43	3	2
3543	167	7370000	HR/RESP MONITOR	18	15	1	0.10	29.47	35	7.69	1.00	42	3	2
15315	167	7370000	HR/RESP MONITOR	18	15	1	0.10	29.47	35	7.69	1.00	42	3	2
17761	167	7370000	HR/RESP MONITOR	18	15	1	0.10	29.47	35	6.69	1.00	41	3	2
18382	574	N100C	PULSE OXIMETER	18	15	1	0.70	4.21	35	7.54	1.00	42	3	2
180476	574	N100C	PULSE OXIMETER	18	15	1	0.70	4.21	35	7.54	1.00	42	3	2
16685	167	7275217	2 CHAN CHART REC	18	15	1	0.42	7.02	37	6.83	1.00	41	3	2

I have reviewed this risk analysis report and have investigated those equipment items for which the risk priority factor has exceeded the average risk for that equipment type. I have taken one of two actions:

1. investigated the equipment item and implemented changes to the maintenance program intended to reduce the risk priority value OR
2. indicated on the printout that the dynamic risk factor program has "overreacted" and the risk factor should be reduced by 5%.

ENGINEER: _____

DATE: _____

FIGURE 3.12 Engineer's risk-assessment report.

manufacturer, model, and equipment type. Assigned values for equipment function and physical risk are constant for each equipment type. The maintenance sensitizing factor enables the clinical engineer to control the algorithm's response to the maintenance level of an entire equipment type. These factors combine to produce the slow risk factor (equipment function+physical risk+maintenance requirements). The unit risk points are multiplied for the unit scaling factor, which allows the clinical engineer to control the algorithm's response to static and dynamic risk components on individual pieces of equipment. This number is then added to the slow risk factor to determine the risk factor for each item. The last two columns are the risk priority that the automated system has assigned and the PM level set by the clinical engineer. This report provides the clinical engineer with information about medical equipment that reflects a higher than normal risk factor for the equipment type to which it belongs.

The proactive risk management tool can be used to individually schedule medical equipment devices for PM based on risk assessment. For example, why should newer patient monitors be maintained at the same maintenance level as older units if the risk can be demonstrated to be less? The tool is used as well to prioritize the planned maintenance

program. For instance, assume a PM cycle every 17 weeks is started on January 1 for a duration of 1 week. Equipment not currently available for PM can be inspected at a later time as a function of the risk priority group for that device. In other words, an equipment item with a risk priority of 2, which is moderately low, would not be overdue for 2/5 of the time between the current and the next PM start date or until the thirteenth week after the start of a PM cycle of 17 weeks. The technicians can complete more critical overdue equipment first and move on to less critical equipment later.

Additionally, since PM is performed with every equipment repair, is it always necessary to perform the following planned PM? Assume for a moment that unscheduled maintenance was performed 10 weeks into the 17 weeks between the two PM periods discussed above. IF the equipment has a higher risk priority of the three, four, or five, the equipment is PMed as scheduled in April. However, if a lower equipment risk priority of one or two is indicated, the planned maintenance is skipped in April and resumed in July. The intent of this application is to reduce maintenance costs, preserve departmental resources, and minimize the war and tear on equipment during testing.

Historically, equipment awaiting service has been placed in the equipment holding area and inspected on a first-in, first-out (FIFO) basis when a technician is available. A client's request to expedite the equipment repair was the singular reason for changing the work priority schedule. The proactive risk-management tool can prioritize the equipment awaiting repair, putting the critical equipment back into service more quickly, subject to the clinical engineer's review.

3.5 Case Studies

Several examples are presented of the proactive risk-assessment tool used to evaluate the performance of medical equipment within a program.

The ventilators in Fig. 3.13 show a decreasing unit risk factor for higher equipment tag numbers. Since devices are put into service with ascending tag numbers and these devices are known to have been purchased over a period of time, the *X* axis represents a chronological progression. The ventilator risk factor is decreasing for newer units and could be attributable to better maintenance technique or manufacturer design improvements. This device is said to have a *time-dependent risk factor*.

A final illustration uses two generations of infusion pumps from the same manufacturer. Figure 3.14 shows the older vintage pump as *Infusion Pump* 1 and the newer version as *Infusion Pump* 2. A linear regression line for the first pump establishes the average risk factor as 53 with a standard deviation of 2.02 for the 93 pumps in the analysis. The second pump, a newer version of the first, had an average risk factor of 50 with a standard deviation of 1.38 for 261 pumps. Both pumps have relatively time-independent risk factors. The proactive risk-management tool reveals that this particular brand of infusion pump in the present maintenance program is stable over time and the newer pump has reduced risk and variability of risk between individual units. Again, this could be attributable to tighter manufacturing control or improvements in the maintenance program.

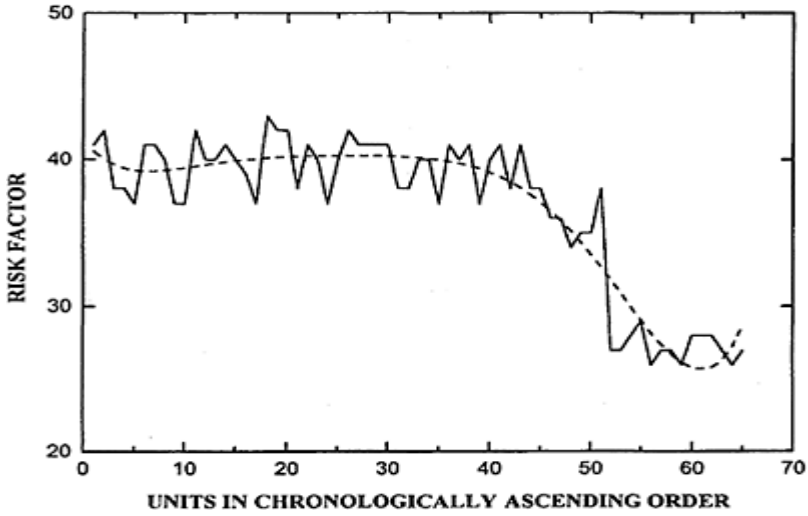


FIGURE 3.13 Ventilator with time-dependent risk characteristics.

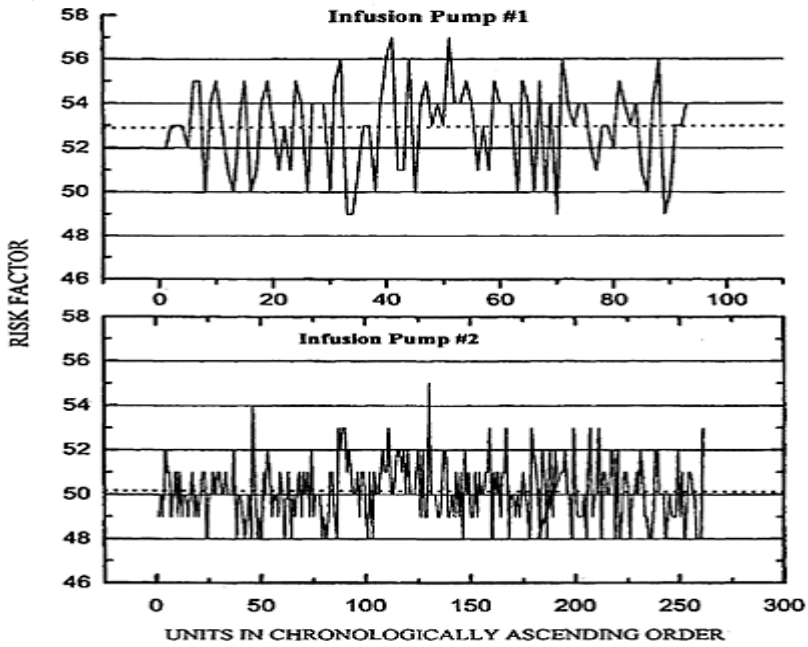


FIGURE 3.14 Time-independent risk characteristics infusion pump #1.

3.6 Conclusions

In summary, superior risk assessment within a medical equipment management program requires better communication, teamwork, and information analysis and distribution among all health care providers. Individually, the clinical engineer cannot provide all the necessary components for managing risk in the health care environment. Using historical information to only address equipment-related problems, after an incident, is not sufficient. The use of a proactive risk-management tool is necessary.

The clinical engineer can use this tool to deploy technical resources in a cost-effective manner. In addition to the direct economic benefits, safety is enhanced as problem equipment is identified and monitored more frequently. The integration of a proactive risk-assessment tool into the equipment management program can more accurately bring to focus technical resources in the health care environment.

References

1. Gullikson ML. 1994. *Biotechnology Procurement and Maintenance II: Technology Risk Management.* Third Annual International Pediatric Colloquium, Houston, Texas.
2. David Y. 1992. *Medical Technology 2001.* Health Care Conference, Texas Society of Certified Public Accountants, San Antonio.
3. Gullikson ML. 1993. *An Automated Risk Management Tool.* Plant, Technology, and Safety Management Series, Joint Commission on the Accreditation of Healthcare Facilities (JCAHO) Monograph 2.
4. Pecht ML, Nash FR. 1994. Predicting the reliability of electronic equipment. *Proc IEEE* 82(7):990.
5. Gullikson ML, David Y. 1993. *Risk-Based Equipment Management Systems.* The 9th National Conference on *Medical Technology Management,* American Society for Hospital Engineering of the American Hospital Association (AHA), New Orleans.
6. Gullikson ML. 1992. *Biomedical Equipment Maintenance System.* 27th Annual Meeting and Exposition, Hospital and Medical Industry Computerized Maintenance Systems, Association for the Advancement of Medical Instrumentation (AAMI), Anaheim, Calif.
7. Fennigkoh L.1989. *Clinical Equipment Management.* Plant, Technology, and Safety Management Monograph 2.
8. David Y, Judd T. 1993. *Medical Technology Management,* Biophysical Measurement Series. Spacelabs Medical, Inc., Redmond, Wash.

4
Clinical Engineering Program Indicators

Dennis D.Autio
Dybonics inc

Robert L.Morris
Dybonics, inc.

The role, organization, and structure of clinical engineering departments in the modern health care environment continue to evolve. During the past 10 years, the rate of change has increased considerably faster than mere evolution due to fundamental changes in the management and organization of health care. Rapid, significant changes in the health care sector are occurring in the United States and in nearly every country. The underlying drive is primarily economic, the recognition that resources are finite.

Indicators are essential for survival of organizations and are absolutely necessary for effective management of change. Clinical engineering departments are not exceptions to this rule. In the past, most clinical engineering departments were task-driven and their existence justified by the tasks performed. Perhaps the most significant change occurring in clinical engineering practice today is the philosophical shift to a more business-oriented, cost-justified, bottom-line-focused approach than has been generally the case in the past.

Changes in the health care delivery system will dictate that clinical engineering departments justify their performance and existence on the same basis as any business, the performance of specific functions at a high-quality level and at a competitive cost. Clinical engineering management philosophy must change from a purely task-driven methodology to one that includes the economics of department performance. Indicators need to be developed to measure this performance. Indicator data will need to be collected and analyzed. The data and indicators must be objective and defensible. If it cannot be measured, it cannot be managed effectively.

Indicators are used to measure performance and function in three major areas. Indicators should be used as internal measurements and monitors of the performance provided by individuals, teams, and the department. These essentially measure what was done and how it was done. Indicators are essential during quality improvement and are used to monitor and improve a process. A third important type of program indicator is the benchmark. It is common knowledge that successful businesses will continue to use benchmarks, even though differing terminology will be used. A business cannot improve its competitive position unless it knows where it stands compared with similar organizations and businesses.

Different indicators may be necessary depending on the end purpose. Some indicators may be able to measure internal operations, quality improvement, and external benchmarks. Others will have a more restricted application.

0-8493-1813-0/03/$0.00+$1.50

It is important to realize that a single indicator is insufficient to provide the information on which to base significant decisions. Multiple indicators are necessary to provide cross-checks and verification. An example might be to look at the profit margin of a business. Even if the profit margin per sale is 100%, the business will not be successful if there are few sales. Looking at single indicators of gross or net profit will correct this deficiency but will not provide sufficient information to point the way to improvements in operations.

4.1 Department Philosophy

A successful clinical engineering department must define its mission, vision, and goals as related to the facility's mission. A mission statement should identify what the clinical engineering department does for the organization. A vision statement identifies the direction and future of the department and must incorporate the vision statement of the parent organization. Department goals are then identified and developed to meet the mission and vision statements for the department and organization. The goals must be specific and attainable. The identification of goals will be incomplete without at least implied indicators. Integrating the mission statement, vision statement, and goals together provides the clinical engineering department management with the direction and constraints necessary for effective planning.

Clinical engineering managers must carefully integrate mission, vision, and goal information to develop a strategic plan for the department. Since available means are always limited, the manager must carefully assess the needs of the organization and available resources, set appropriate priorities, and determine available options. The scope of specific clinical engineering services to be provided can include maintenance, equipment management, and technology management activities. Once the scope of services is defined, strategies can be developed for implementation. Appropriate program indicators must then be developed to document, monitor, and manage the services to be provided. Once effective indicators are implemented, they can be used to monitor internal operations and quality-improvement processes and complete comparisons with external organizations.

Monitoring Internal Operations

Indicators may be used to provide an objective, accurate measurement of the different services provided in the department. These can measure specific individual, team, and departmental performance parameters. Typical indicators might include simple tallies of the quantity or level of effort for each activity, productivity (quantity/effort), percentage of time spent performing each activity, percentage of scheduled IPMs (inspection and preventive maintenance procedures) completed within the scheduled period, mean time per job by activity, repair jobs not completed within 30 days, parts order for greater than 60 days, etc.

Process for Quality Improvement

When program indicators are used in a quality-improvement process, an additional step is required. Expectations must be quantified in terms of the indicators used. Quantified expectations result in the establishment of a threshold value for the indicator that will precipitate further analysis of the process. Indicators combined with expectations (threshold values of the indicators) identify the opportunities for program improvement. Periodic monitoring to determine if a program indicator is below (or above, depending on whether you are measuring successes or failures) the established threshold will provide a flag to whether the process or performance is within acceptable limits. If it is outside acceptable limits for the indicator, a problem has been identified. Further analysis may be required to better define the problem. Possible program indicators for quality improvement might include the number of repairs completed within 24 or 48 hours, the number of callbacks for repairs, the number of repair problems caused by user error, the percentage of hazard notifications reviewed and acted on within a given time frame, meeting time targets for generating specification, evaluation or acceptance of new equipment, etc.

An example might be a weekly status update of the percentage of scheduled IPMs completed. Assume that the department has implemented a process in which a group of scheduled IPMs must be completed within 8 weeks. The expectation is that 12% of the scheduled IPMs will be completed each week. The indicator is the percentage of IPMs completed. The threshold value of the indicator is 12% per week increase in the percentage of IPMs completed. To monitor this, the number of IPMs that were completed must be tallied, divided by the total number scheduled, and multiplied by 100 to determine the percentage completed. If the number of completed IPMs is less than projected, then further analysis would be required to identify the source of the problem and determine solutions to correct it. If the percentage of completed IPMs were equal to or greater than the threshold or target, then no action would be required.

External Comparisons

Much important and useful information can be obtained be carefully comparing one clinical engineering program with others. This type of comparison is highly valued by most hospital administrators. It can be helpful in determining performance relative to competitors. External indicators or benchmarks can identify specific areas of activity in need of improvement. They offer insights when consideration is being given to expanding into new areas of support. Great care must be taken when comparing services provided by clinical engineering departments located in different facilities. There are a number of factors that must be included in making such comparisons; otherwise, the results can be misleading or misinterpreted. It is important that the definition of the specific indicators used be well understood, and great care must be taken to ensure that the comparison utilizes comparable information before interpreting the comparisons. Failure to understand the details and nature of the comparison and just using the numbers directly will likely result in inappropriate actions by managers and administrators. The process of analysis and explanation of differences in benchmark values between a clinical engineering department and a competitor (often referred to as *gap analysis*) can lead to increased insight into department operations and target areas for improvements.

Possible external indicators could be the labor cost per hour, the labor cost per repair, the total cost per repair, the cost per bed supported, the number of devices per bed supported, percentage of time devoted to repairs versus IPMs versus consultation, cost of support as a percentage of the acquisition value of capital inventory, etc.

4.2 Standard Database

In God we trust…all others bring data!
Florida Power and Light

Evaluation of indicators requires the collection, storage, and analysis of data from which the indicators can be derived. A standard set of data elements must be defined. Fortunately, one only has to look at commercially available equipment management systems to determine the most common data elements used. Indeed, most of the high-end software systems have more data elements than many clinical engineering departments are willing to collect. These standard data elements must be carefully defined and understood. This is especially important if the data will later be used for comparisons with other organizations. Different departments often have different definitions for the same data element. It is crucial that the data collected be accurate and complete. The members of the clinical engineering department must be trained to properly gather, document, and enter the data into the database. It makes no conceptual difference if the database is maintained on paper or using computers. Computers and their databases are ubiquitous and so much easier to use that usually more data elements are collected when computerized systems are used. The effort required for analysis is less and the level of sophistication of the analytical tools that can be used is higher with computerized systems.

The clinical engineering department must consistently gather and enter data into the database. The database becomes the practical definition of the services and work performed by the department. This standardized database allows rapid, retrospective analysis of the data to determine specific indicators identifying problems and assist in developing solutions for implementation. A minimum database should allow the gathering and storage of the following data:

In-House Labor. This consists of three elements: the number of hours spent providing a particular service, the associated labor rate, and the identity of the individual providing the service. The labor cost is not the hourly rate the technician is paid multiplied by the number of hours spent performing the service. It should include the associated indirect costs, such as benefits, space, utilities, test equipment, and tools, along with training, administrative overhead, and many other hidden costs. A simple, straightforward approach to determine an hourly labor rate for a department is to take the total budget of the department and subtract parts' costs, service contract costs, and amounts paid to outside vendors. Divide the resulting amount by the total hours spent providing services as determined from the database. This will provide an average hourly rate for the department.

Vendor Labor. This should include hours spent and rate, travel, and zone charges, and any perdiem costs associated with the vendor supplied service.
Parts. Complete information on parts is important for any retrospective study of services provided. This information is similar for both in-house and vendor-provided service. It should include the part number, a description of the part, and its cost, including any shipping.
Time. It is important to include a number of time stamps in the data. These should include the date the request was received, data assigned, and date completed.
Problem Identification. Both a code for rapid computer searching and classification and a free text comment identifying the nature of the problem and description of service provided are important. The number of codes should be kept to as few as possible. Detailed classification schemes usually end up with significant inaccuracies due to differing interpretations of the fine gradations in classifications.
Equipment Identification. Developing an accurate equipment history depends on reliable means of identifying the equipment. This usually includes a department- and/or facility-assigned unique identification number as well as the manufacturer, vendor, model, and serial number. Identification numbers provided by asset management are often inadequate to allow tracking of interchangeable modules or important items with a value less than a given amount. Acquisition cost is a useful data element.
Service Requester. The database should include elements allowing identification of the department, person, telephone number, cost center, and location of the service requester.

4.3 Measurement Indicators

Clinical engineering departments must gather objective, quantifiable data in order to assess ongoing performance, identify new quality-improvement opportunities, and monitor the effect of improvement action plans. Since resources are limited and everything cannot be measured, certain selection criteria must be implemented to identify the most significant opportunities for indicators. High-volume, high-risk, or problem-prone processes require frequent monitoring of indicators. A new indicator may be developed after analysis of ongoing measurements or feedback from other processes. Customer feedback and surveys often can provide information leading to the development of new indicators. Department management, in consultation with the quality-management department, typically determines what indicators will be monitored on an ongoing basis. The indicators and resulting analysis are fed back to individuals and work teams for review and improvement of their daily work activities. Teams may develop new indicators during their analysis and implementation of solutions to quality-improvement opportunities.

An *indicator* is an objective, quantitative measurement of an outcome or process that relates to performance quality. The event being assessed can be either desirable or undesirable. It is objective in that the same measurement can be obtained by different observers. This indicator represents quantitative, measured data that are gathered for further analysis. Indicators can assess many different aspects of quality, including accessibility, appropriateness, continuity, customer satisfaction, effectiveness, efficacy, efficiency, safety, and timeliness.

A program indicator has attributes that determine its utility as a performance measure. The reliability and variability of the indicator are distinct but related characteristics. An indicator is reliable if the same measurement can be obtained by different observers. A valid indicator is one that can identify opportunities for quality improvement. As indicators evolve, their reliability and validity should improve to the highest level possible.

An indicator can specify a part of a process to be measured or the outcome of that process. An outcome indicator assesses the results of a process. Examples include the percentage of uncompleted, scheduled IPMs, or the number of uncompleted equipment repairs not completed within 30 days. A process indicator assesses an important and discrete activity that is carried out during the process. An example would be the number of anesthesia machines in which the scheduled IPM failed or the number of equipment repairs awaiting parts that are uncompleted within 30 days.

Indicators also can be classified as sentinel event indicators and aggregate data indicators. A performance measurement of an individual event that triggers further analysis is called a *sentinel-event indicator*. These are often undesirable events that do not occur often. These are often related to safety issues and do not lend themselves easily to quality-improvement opportunities. An example may include equipment failures that result in a patient injury.

An aggregate data indicator is a performance measurement based on collecting data involving many events. These events occur frequently and can be presented as a continuous variable indicator or as rate-based indicators. A continuous variable indicator is a measurement where the value can fall anywhere along a continuous scale. Examples could be the number of IPMs scheduled during a particular month or the number of repair requests received during a week. A rate-based variable indicator is the value of a measurement that is expressed as a proportion or a ratio. Examples could be the percentage of IPMs completed each month or the percentage of repairs completed within one workday.

General indicators should be developed to provide a baseline monitoring of the department's performance. They also should provide a cross-check for other indicators. These indicators can be developed to respond to a perceived need within a department or to solve a specific problem.

4.4 Indicator Management Process

The process to develop, monitor, analyze and manage indicators is shown in Fig. 4.1. The different steps in this process include defining the indicator, establishing the threshold, monitoring the indicator, evaluating the indicator, identifying quality-improvement opportunities, and implementing action plans.

Define Indicator. The definition of the indicator to be monitored must be carefully developed. This process includes at least five steps. The event or outcome to be measured must be described. Define any specific terms that are used. Categorize the indicator (sentinel event or rate-based, process or outcome, desirable or undesirable). The purpose for this indicator must be defined, as well as how it is used in specifying and assessing the particular process or outcome.

Establish Threshold. A threshold is a specific data point that identifies the need for the department to respond to the indicator to determine why the threshold was reached. Sentinel-event indicator thresholds are set at zero. Rate-indicator thresholds are more complex to define because they may require expert consensus or definition of the department's objectives. Thresholds must be identified, including the process used to set the specific level.

Monitor Indicator. Once the indicator is defined, the data-acquisition process identifies the data sources and data elements. As these data are gathered, they must be validated for accuracy and completeness. Multiple indicators can be used for data validation and cross-checking. The use of a computerized database allows rapid access to the data. A database management tool allows quick sorting and organization of the data. Once gathered, the data must be presented in a format suitable for evaluation. Graphic presentation of data allows rapid visual analysis for thresholds, trends, and patterns.

Evaluate Indicator. The evaluation process analyzes and reports the information. This process includes comparing the information with established thresholds and analyzing for any trends or patterns. A *trend* is the general direction the indicator measurement takes over a period of time and may be desirable or undesirable. A *pattern* is a grouping or distribution of indicator measurements. A pattern analysis is often triggered when thresholds are crossed or trends identified. Additional indicator information is often required. If an indictor threshold has not been reached, no further action may be necessary, other than continuing to monitor this indicator. The department also may decide to improve its performance level by changing the threshold.

Factors may be present leading to variation of the indicator data. These factors may include failure of the technology to perform properly, failure of the operators to use the technology properly, and failure of the organization to provide the necessary resources to implement this technology properly. Further analysis of these factors may lead to quality-improvement activities later.

Identify Quality-Improvement Opportunity. A quality-improvement opportunity may present itself if an indicator threshold is reached, a trend is identified, or a pattern is recognized. Additional information is then needed to further define the process and improvement opportunities. The first step in the process is to identify a team. This team must be given the necessary resources to complete this project, a timetable to be followed, and an opportunity to periodically update management on the status of the project. The initial phase of the project will analyze the process and establish the scope and definition of the problem. Once the problem is defined, possible solutions can be identified and analyzed for potential implementation. A specific solution to the problem is then selected. The solution may include modifying existing indictors or thresholds to more appropriate values, modifying steps to improve existing processes, or establishing new goals for the department.

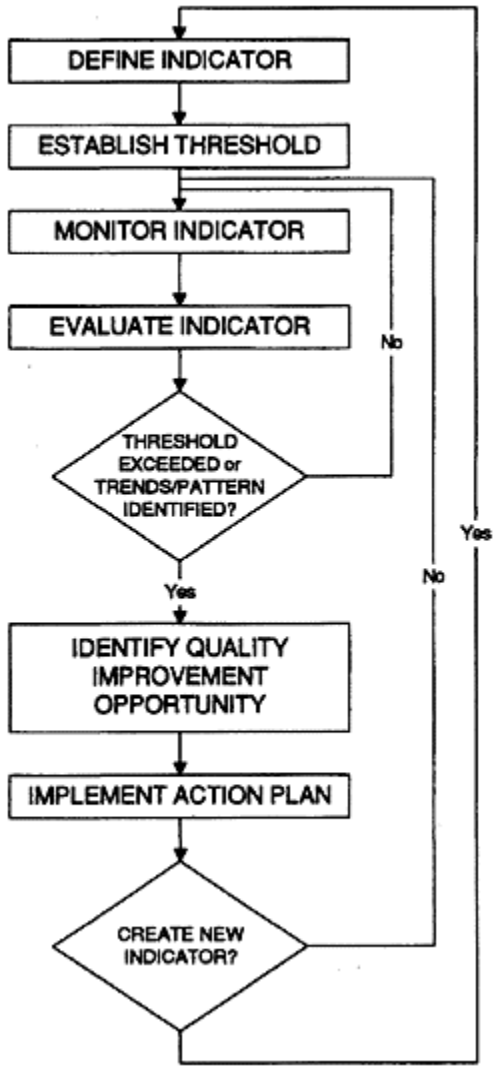


FIGURE 4.1 Indicator management process.

Implement Action Plan. An action plan is necessary to identify how the quality-improvement solution will be implemented. This includes defining the different tasks to be performed, the order in which they will be addressed, who will perform each task, and how this improvement will be monitored. Appropriate resources must again be identified and a timetable developed prior to implementation. Once the action plan is implemented, the indicators are monitored and evaluated to verify appropriate changes in the process. New indicators and thresholds may need to be developed to monitor the solution.

4.5 Indicator Example 1: Productivity Monitors

Defines Indicators. Monitor the productivity of technical personnel, teams, and the department. Productivity is defined as the total number of documented service support hours compared with the total number of hours available. This is a desirable rate-based outcome indicator. Provide feedback to technical staff and hospital administration regarding utilization of available time for department support activities.
Establish Thresholds. At least 50% of available technician time will be spent providing equipment maintenance support services (revolving equipment problems and scheduled IPMs). At least 25% of available technician time will be spent providing equipment management support services (installations, acceptance testing, incoming inspections, equipment inventory database management, hazard notification review).
Monitor Indicator. Data will be gathered every 4 weeks from the equipment work-order history database. A trend analysis will be performed with data available from previously monitored 4-week intervals. These data will consist of hours worked on completed and uncompleted jobs during the past 4-week interval.

Technical staff available hours is calculated for the 4-week interval. The base time available is 160 hours (40 hours/week×4 weeks) per individual. Add to this any overtime worked during the interval. Then subtract any holidays, sick days, and vacation days within the interval.

CJHOURS: Hours worked on completed jobs during the interval
UJHOURS: Hours worked on uncompleted jobs during the interval
AHOURS: Total hours available during the 4-week interval

Productivity=(CJHOURS+UJHOURS)/AHOURS

Evaluate Indicator. The indicator will be compared with the threshold, and the information will be provided to the individual. The individual team member data can be summed for team review. The data from multiple teams can be summed and reviewed by the department. Historical indicator information will be utilized to determine trends and patterns.
Quality-Improvement Process. If the threshold is not met, a trend is identified, or a pattern is observed, a quality-improvement opportunity exists. A team could be formed to review the indicator, examine the process that the indicator measured, define the problem encountered, identify ways to solve the problem, and select a solution. An action plan will then be developed to implement this solution.
Implement Action Plan. During implementation of the action plan, appropriate indicators will be used to monitored the effectiveness of the action plan.

4.6 Indicator Example 2: Patient Monitors IPM Completion Time

Define Indicator. Compare the mean to complete an IPM for different models of patient monitors. Different manufacturers of patient monitors have different IPM requirements. Identify the most timely process to support this equipment.
Establish Threshold. The difference between the mean time to complete an IMP for different models of patient monitors will not be greater than 30% of the lessor time.

Monitor Indicator. Determine the mean time to complete an IPM for each model of patient monitor. Calculate the percentage difference between the mean time for each model and the model with the least mean time.
Evaluate Indicator. The mean time to complete IPMs was compared between the patient monitors, and the maximum difference noted was 46%. A pattern also was identified in which all IPMs for that one particular monitor averaged 15 minutes longer than those of other vendors.
Quality-Improvement Process. A team was formed to address this problem. Analysis of individual IPM procedures revealed that manufacturer X requires the case to be removed to access internal filters. Performing an IPM for each monitor required moving and replacing 15 screws for each of the 46 monitors. The team evaluated this process and identified that 5 minutes could be saved from each IPM if an electric screwdriver was utilized.
Implement Action Plan. Electric screwdrivers were purchased and provided for use by the technician. The completion of one IPM cycle for the 46 monitors would pay for two electric screwdrivers and provide 4 hours of productive time for additional work. Actual savings were greater because this equipment could be used in the course of daily work.

4.7 Summary

In the ever-changing world of health care, clinical engineering departments are frequently being evaluated based on their contribution to the corporate bottom line. For many departments, this will require difficult and painful changes in management philosophy. Administrators are demanding quantitative measures of performance and value. To provide the appropriate quantitative documentation required by corporate managers, a clinical engineering a manager must collect available data that are reliable and accurate. Without such data, analysis is valueless. Indicators are the first step in reducing the data to meaningful information that can be easily monitored and analyzed. The indicators can then be used to determine department performance and identify opportunities for quality improvement.

Program indicators have been used for many years. What must change for clinical engineering departments is a conscious evaluation and systematic use of indicators. One traditional indicator of clinical engineering department success is whether the department's budget is approved or not. Unfortunately, approval of the budge as an indicator, while valuable, does not address the issue of predicting long-term survival, measuring program and quality improvements, or allowing frequent evaluation and changes.

There should be monitored indicators for every significant operational aspect of the department. Common areas where program indicators can be applied include monitoring interval department activities, quality-improvement processes, and benchmarking. Initially, simple indicators should be developed. The complexity and number of indicators should change as experience and needs demand.

The use of program indicators is absolutely essential if a clinical engineering department is to survive. Program and survival are now determined by the contribution of the department to the bottom line of the parent organization. Indicators must be

developed and utilized to determine the current contribution of the clinical engineering department to the organization. Effective utilization and management of program indicators will ensure future department contributions.

References

AAMI. 1993. *Management Information Report MIR 1: Design of Clinical Engineering Quality Assurance Risk Management Programs.* Arlington, Va, Association for the Advancement of Medical Instrumentation.

AAMI. 1993. *Management Information Report MIR 2: Guideline for Establishing and Administering Medical Instrumentation Maintenance Programs.* Arlington, Va, Association for the Advancement of Medical Instrumentation.

AAMI. 1994. *Management Information Report MIR 3: Computerized Maintenance Management Systems for Clinical Engineering.* Arlington, Va, Association for the Advancement of Medical Instrumentation.

Bauld TJ. 1987. Productivity: Standard terminology and definitions. *J Clin Eng* 12(2): 139.

Betts WE 1989. Using productivity measures in clinical engineering departments. *Biomed Instrum Technol* 23(2):120.

Bronzino JD. 1992. *Management of Medical Technology: A Primer for Clinical Engineers.* Stoneham, Mass, Butterworth-Heinemann.

Coopers and Lybrand International, AFSM. 1994. *Benchmarking Impacting the Boston Line.* Fort Myers, Fla, Association for Services Management International.

David Y, Judd TM. 1993. Risk management and quality improvement. In *Medical Technology Management,* pp 72–75. Redmond, Wash, SpaceLab Medical.

David Y, Rohe D. 1986. Clinical engineering program productivity and measurement. *J Clin Eng* 11(6):435.

Downs KJ, McKinney WD. 1991. Clinical engineering workload analysis: A proposal for standardization. *Biomed Instrum Technol* 25(2):101.

Fennigkoh L. 1986. *ASHE Technical Document No 055880: Medical Equipment Maintenance Performance Measures.* Chicago, American Society for Hospital Engineers.

Furst E. 1986. Productivity and cost-effectiveness of clinical engineering. *J Clin Eng* 11(2): 105.

Gordon GJ. 1995. *Breakthrough Management—A New Model For Hospital Technical Services.* Arlington, Va, Association for the Advancement of Medical Instrumentation.

Hertz E. 1990. Developing quality indicators for a clinical engineering department. In *Plant, Technology and Safety Management Series: Measuring Quality in PTSM.* Chicago, Joint Commission on Accreditation of Healthcare Organizations.

JCAHO. 1990. *Primer on Indicator Development and Application, Measuring Quality in Health Care.* Oakbrook, 111, Joint Commission on Accreditation of Healthcare Organizations.

JCAHO. 1994. *Framework for Improving Performance.* Oakbrook, 111, Joint Commission on Accreditation of Healthcare Organizations.

Keil OR. 1989. The challenge of building quality into clinical engineering programs. *Biomed Instrum Technol* 23(5):354.

Lodge DA. 1991. Productivity, efficiency, and effectiveness in the management of healthcare technology: An incentive pay proposal. *J Clin Eng* 16(1):29.

Mahachek AR. 1987. Management and control of clinical engineering productivity: A case study. *J Clin Eng* 12(2):127.

Mahachek AR. 1989. Productivity measurement. Taking the first steps. *Biomed Instrum Technol* 23:16.

Selsky DB, Bell DS, Benson D, et al. 1991. Biomedical equipment information management for the next generation. *Biomed Instrum Technol* 25(1):24.

Sherwood MK. 1991. Quality assurance in Biomedical or clinical engineering. *J Clin Eng* 16(6):479.
Stiefel RH. 1991. Creating a quality measurement system for clinical engineering. *Biomed Instrum Technol.* 25(1):17.

5
Quality of Improvement and Team Building

Joseph P.McClain
Walter Reed Army
Medical Center

In today's complex health care environment, quality improvement and team building must go hand in hand. This is especially true for Clinical Engineers and Biomedical Equipment Technicians as the diversity of the field increases and technology moves so rapidly that no one can know all that needs to be known without the help of others. Therefore, it is important that we work together to ensure quality improvement. Ken Blachard, the author of the *One Minute Manager* series, has made the statement that "all of us are smarter than any one of us"—a synergy that evolves from working together. Throughout this chapter we will look closely at defining quality and the methods for continuously improving quality, such as collecting data, interpreting indicators, and team building. All this will be put together, enabling us to make decisions based on scientific deciphering of indicators.

Quality is defined as conformance to customer or user requirements. If a product or service does what it is supposed to do, it is said to have high quality. If the product or service fails its mission, it is said to be of low quality. Dr. W.Edward Demings, who is known to many as the "father of quality," defined it as surpassing customer needs and expectations throughout the life of the product or service.

Dr. Demings, a trained statistician by profession, formed his theories on quality during World War II while teaching industry how to use statistical methods to improve the quality of military production. After the war, he focused on meeting customer or consumer needs and acted as a consultant to lapanese organizations to change consumers' perceptions that "Made in lapan" meant junk. Dr. Demings predicted that people would be demanding lapanese products in just 5 years, if they used his methods. However, it only took 4, and the rest is history.

5.1 Deming's 14 Points

1. Create constancy of purpose toward improvement of product and service, with an aim to become competitive and to stay in business and provide jobs.
2. Adopt the new philosophy. We are in a new economic age. Western management must awaken and lead for change.
3. Cease dependence on inspection to achieve quality. Eliminate the needs for mass inspection by first building in quality.
4. Improve constantly and forever the system of production and service to improve quality and productivity and thus constantly decrease costs.
5. Institute training on the job.

0-8493-1813-0/03/$0.00+$1.50

6. Institute leadership: The goal is to help people, machines, and gadgets to do a better job.
7. Drive out fear so that everyone may work effectively for the organization.
8. Break down barriers between departments
9. Eliminate slogans, exhortations, and targets for the workforce.
10. Eliminate work standards (quota) on the factory floor.
11. Substitute leadership: Eliminate management by objective, by numbers, and numerical goals.
12. Remove barriers that rob the hourly worker of the right to pride of workmanship.
13. Institute a vigorous program of education and self-improvement.
14. Encourage everyone in the company to work toward accomplishing transformation. Transformation is everyone's job.

5.2 Zero Defects

Another well-known quality theory, called *zero defects (ZD)*, was established by Philip Crosby. It got results for a variety of reasons. The main reasons are as follows:

1. A strict and specific management standard. Management, including the supervisory staff, do not use vague phrases to explain what it wants. It made the quality standard very clear: *Do it the right way from the start.* As Philip Crosby said, "What standard would you set on how many babies nurses are allowed to drop?"
2. Complete commitment from everyone. Interestingly, Crosby denies that ZD was a motivational program. But ZD worked because everyone got deeply into the act. Everyone was encouraged to spot problems, detect errors, and prescribe ways and means for their removal. This commitment is best illustrated by the ZD pledge: "I freely pledge myself to make a constant, conscious effort to do my job right the first time, recognizing that my individual contribution is a vital part of the overall effort."
3. Removal of actions and conditions that cause errors. Philip Crosby claimed that at ITT, where he was vice-president for quality, 90% of all error causes could be acted on and fully removed by first-line supervision. In other words, top management must do its part to improve conditions, but supervisors and employees should handle problems directly. Errors, malfunctions, and/or variances can best be corrected where the rubber hits the road—at the source.

5.3 TQM (Total Quality Management)

The most recent quality theory that has found fame is called TQM *(Total Quality Management)*. It is a strategic, integrated management system for achieving customer satisfaction that involves all managers and employees and uses quantitative methods to continuously improve an organization's processes. *Total Quality Management* is a term coined in 1985 by the Naval Air Systems Command to describe its management approach to quality improvement. Simply put, Total Quality Management is a management approach to long-term success through customer satisfaction. Total Quality Management

includes the following three principles: (1) achieving customer satisfaction, (2) making continuous improvement, and (3) giving everyone responsibility. TQM includes eight practices. These practices are (1) focus on the customer, (2) effective and renewed communications, (3) reliance on standards and measures, (4) commitment to training, (5) top management support and direction, (6) employee involvement, (7) rewards and recognition, and (8) long-term commitment.

5.4 CQI (Continuous Quality Improvement)

Step 8 of the total quality management practices leads us to the quality concept coined by the Joint Commission on Accreditation of Healthcare Organizations and widely used by most health care agencies. It is called CQI *(Continuous Quality Management)*. The principles of CQI are as follows:

Unity of Purpose

- Unity is established throughout the organization with a clear and widely understood vision.
- Environment nurtures total commitment from all employees.
- Rewards go beyond benefits and salaries to the belief that *"We are family"* and *"We do excellent work."*

Looking for Faults in the Systems

- Eighty percent of an organization's failures are the fault of management-controlled systems.
- Workers can control fewer than 20% of the problems.
- Focus on rigorous improvement of every system, and cease blaming individuals for problems (the 80/20 rule of J.M.Juran and the 19th-century economist Vilfredo Pareto).

Customer Focus

- Start with the customer.
- The goal is to meet or exceed customer needs and give lasting value to the customer.
- Positive returns will follow as customers boast of the company's quality and service.

Obsession with Quality

- Everyone's job.
- Quality is relentlessly pursued through products and services that delight the customer.
- Efficient and effective methods of execution.

Recognizing the Structure in Work

- All work has structure.
- Structure may be hidden behind workflow inefficiency.
- Structure can be studied, measure, analyzed, and improved.

Freedom through Control

- There is control, yet freedom exists by eliminating micromanagement.
- Employees standardize processes and communicate the benefits of standardization.
- Employees reduce variation in the way work is done.
- Freedom comes as changes occur, resulting in time to spend on developing improved processes, discovering new markets, and adding other methods to increase productivity.

Continued Education and Training

- Everyone is constantly learning.
- Educational opportunities are made available to employees.
- Greater job mastery is gained and capabilities are broadened.

Philosophical Issues on Training

- Training must stay tuned to current technology.
- Funding must be made available to ensure that proper training can be attained.
- Test, measurement, and diagnostic equipment germane to the mission must be procured and technicians trained on its proper use, calibration, and service.
- Creativity must be used to obtain training when funding is scarce.
    - Include training in equipment procurement process.
    - Contact manufacturer or education facility to bring training to the institution.
    - Use local facilities to acquire training, thus eliminating travel cost.
    - Allow employees to attend professional seminars where a multitude of training is available.

Teamwork

- Old rivalries and distrust are eliminated.
- Barriers are overcome.
- Teamwork, commitment to the team concept, and partnerships are the focus.
- Employee empowerment is critical in the CQI philosophy and means that employees have the authority to make well-reasoned, data-based decisions. In essence, they are entrusted with the legal power to change processes through a rational, scientific approach.

Continuous quality improvement is a means for tapping knowledge and creativity, applying participative problem solving, finding and eliminating problems that prevent quality, eliminating waste, instilling pride, and increasing teamwork. Further it is a means

for creating an atmosphere of innovation for continued and permanent quality improvement. Continuous quality improvement as outlined by the Joint Commission on Accreditation of Healthcare Organizations is designed to improve the work processes within and across organizations.

5.5 Tools Used for Quality Improvement

The tools listed on the following pages will assist in developing quality programs, collecting data, and assessing performance indicators within the organization. These tools include several of the most frequently used and most of the **seven tools of quality**. The seven tools of quality are tools that help health care organizations understand their processes in order to improve them. The tools are the cause-and-effect diagram, check sheet, control chart, flowchart, histogram, Pareto chart, and scatter diagram. Additional tools shown are the Shewhart cycle (PDCA process) and the bar chart. The Clinical Engineering Manager must access the situation and determine which tool will work best for his or her situational needs.

Two of the seven tools of quality discussed above are not illustrated. These are the scatter diagram and the check sheet. The scatter diagram is a graphic technique to analyze the relationship between two variations and the check sheet is simple data-recording device. The check sheet is custom designed by the user, which facilitates interpretation of the results. Most Biomedical Equipment Technicians use the check sheet on a daily basis when performing preventive maintenance, calibration, or electrical safety checks.

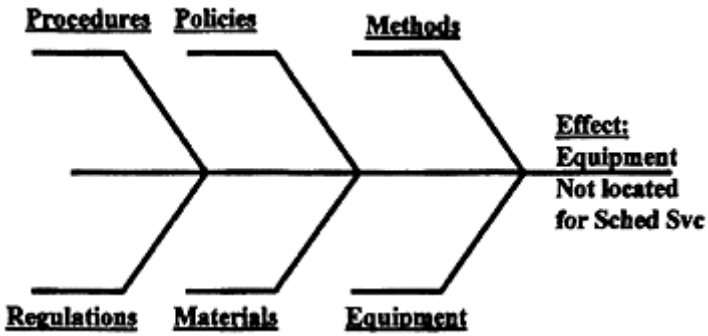


FIGURE 5.1 Cause-and-effect or Ishikawa chart.

Cause-and-Effect or Ishikawa Chart

This is a tool for analyzing process dispersion (Fig. 5.1). The process was developed by Dr. Karou Ishikawa and is also known as the *fishbone diagram* because the diagram resembles a fish skeleton. The diagram illustrates the main causes and subcauses leading to an effect. The cause-and-effect diagram is one of the seven tools of quality.

The following is an overview of the process:

- Used in group problem solving as a *brainstorming tool* to explore and display the possible causes of a particular problem.
- The *effect* (problem, concern, or opportunity) that is being investigated is stated on the right side, while the contributing *causes* are grouped in component categories through group brainstorming on the left side.
- This is an extremely effective tool for focusing a group brainstorming session.
- Basic components include environment, methods (measurement), people, money information, materials, supplies, capital equipment, and intangibles.

Control Chart

A control chart is a graphic representation of a characteristic of a process showing plotted values of some statistic gathered from that characteristic and one or two control limits (Fig. 5.2). It has two basic uses:

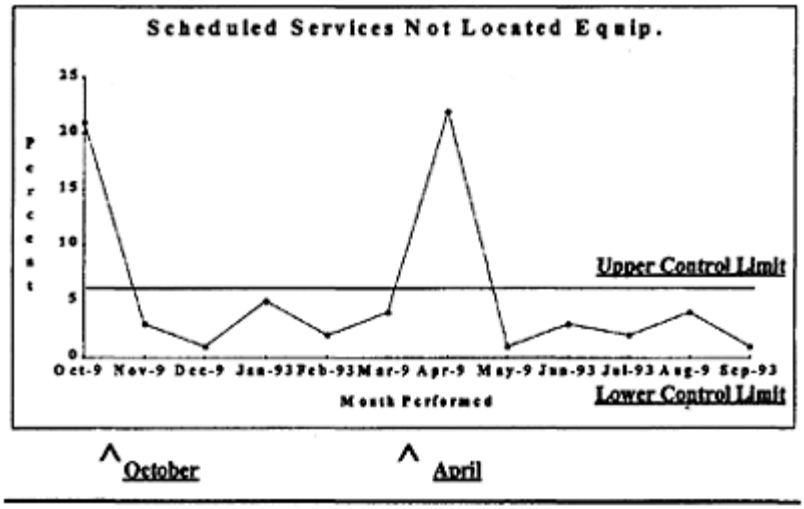


FIGURE 5.2 Control chart.

- as a judgment to determine if the process is in control
- as an aid in achieving and maintaining statistical control

(This chart was used by Dr. W.A.Shewhart for a continuing test of statistical significance.) A control chart is a chart with a baseline, frequently in time order, on which measurement or counts are represented by points that are connected by a straight line with an upper and lower limit. The control chart is one of the seven tools of quality.

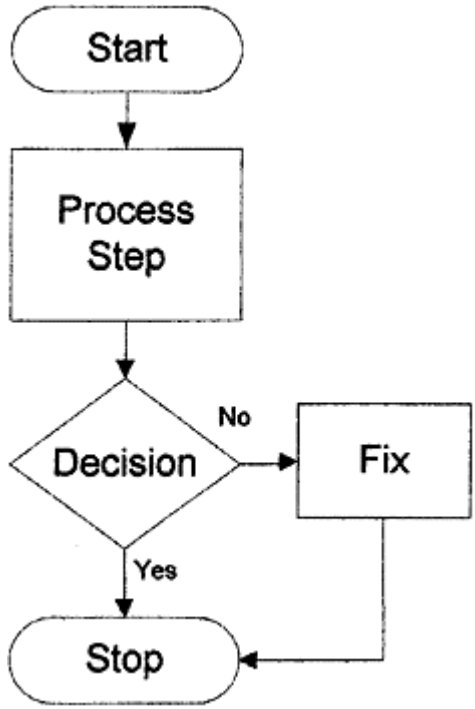


FIGURE 5.3 Flowchart.

Flowchart

A flowchart is a pictorial representation showing all the steps of a process (Fig. 5.3). Flowcharts provide excellent documentation of a program and can be a useful tool for examining how various steps in a process are related to each other. Flowcharting uses easily recognizable symbols to represent the type of processing performed. The flowchart is one of the seven tools of quality.

Histogram

A graphic summary of variation in a set of data is a histogram (Fig. 5.4). The pictorial nature of the histogram lets people see patterns that are difficult to see in a simple table of numbers. The histogram is one of the seven tools of quality.

Pareto Chart

A Pareto chart is a special form of vertical bar graph that helps us to determine which problems to solve and in what order (Fig. 5.5). It is based on the Pareto principle, which was first developed by J.M.Juran in 1950. The principle, named after the 19th-century

economist Vilfredo Pareto, suggests that most effects come from relatively few causes; i.e., 80% of the effects come from 20% of the possible causes.

Doing a Pareto chart, based on either check sheets or other forms of data collection, helps us direct our attention and efforts to truly important problems. We will generally gain more by working on the tallest bar than tackling the smaller bars. The Pareto chart is one of the seven tools of quality.

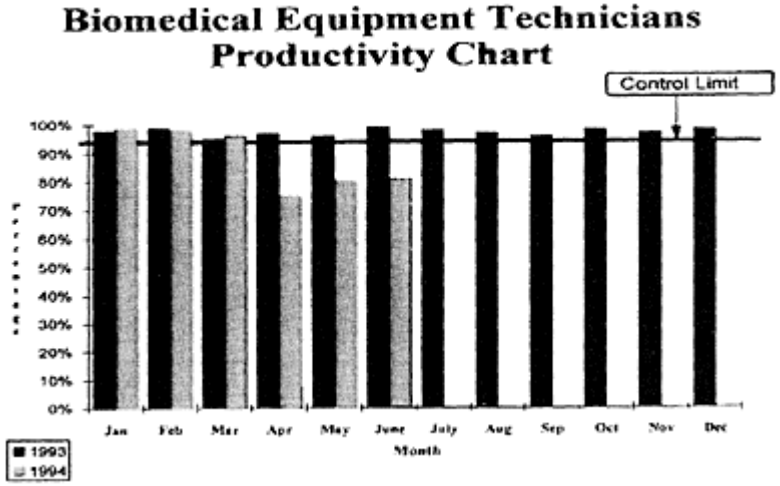


FIGURE 5.4 Histogram.

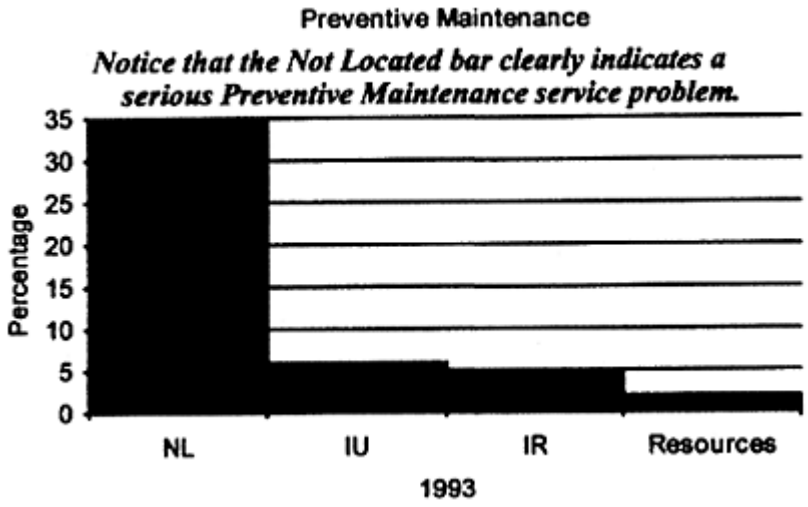


FIGURE 5.5 Pareto chart.

- **Step 1 - Plan (P): Collect data upon which a plan can be constructed.**
- **Step 2 - Do (D): Take the necessary actions that further the developed plan.**
- **Step 3 - Check (C): Check the results of our actions by collecting data to make sure that we have achieved what we planned.**
- **Step 4 - Act (A): Act by making necessary changes to achieve Customer Satisfaction.**

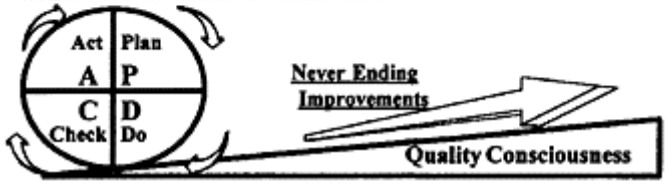


FIGURE 5.6 The Shewhart cycle.

The Plan-Do-Check-Act or Shewhart Cycle

This is a four-step process for quality improvement that is sometimes referred to as the *Deming cycle* (Fig. 5.6). One of the consistent requirements of the cycle is the long-term commitment required. The Shewart cycle or PDCA cycle is outlined here and has had overwhelming success when used properly. It is also a very handy tool to use in understanding the quality cycle process. The results of the cycle are studied to determine what was learned, what can be predicted, and appropriate changes to be implemented.

5.6 Quality Performance Indicators (QPI)

An *indicator* is something that suggests the existence of a fact, condition, or quality—an omen (a sign of future good or evil.) It can be considered as evidence of a manifestation or symptom of an incipient failure or problem. Therefore, quality performance indicators are measurements that can be used to ensure that quality performance is continuous and will allow us to know when incipient failures are starting so that we may take corrective and preventive actions.

QPI analysis is a five-step process:

Step 1: Decide what performance we need to track.

Step 2: Decide the data that need to be collected to track this performance.

Step 3: Collect the data.

Step 4: Establish limits, a parameter, or control points.

Step 5: Utilize BME (management by exception)—where a performance exceeds the established control limits, it is indicating a quality performance failure, and corrective action must be taken to correct the problem.

In the preceding section, there were several examples of QPIs. In the Pareto chart, the NL=not located, IU=in use, IR=in repair. The chart indicates that during the year 1994, 35% of the equipment could not be located to perform preventive maintenance services. This indicator tells us that we could eventually have a serious safety problem that could impact on patient care, and if not corrected, it could prevent the health care facility from meeting accreditation requirements. In the control chart example, an upper control limit of 6% "not located equipment" is established as acceptable in any one month. However, this upper control limit is exceeded during the months of April and October. This QPI could assist the clinical and/or Biomedical Equipment Manager in narrowing the problem down to a 2-month period. The histogram example established a lower control limit for productivity at 93%. However, productivity started to drop off in May, June, and July. This QPI tells the manager that something has happened that is jeopardizing the performance of his or her organization. Other performance indicators have been established graphically in Figs. 5.7 and 5.8. See if you can determine what the indicators are and what the possible cause might be. You may wish to use these tools to establish QPI tracking germane to your own organization.

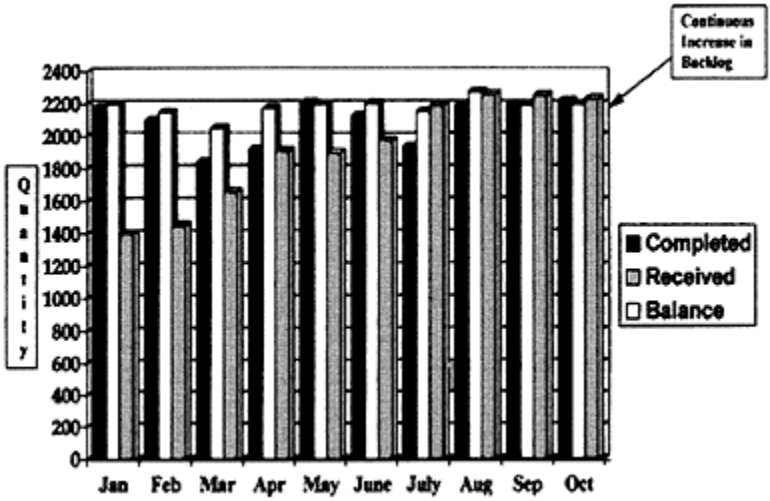


FIGURE 5.7 Sample repair service report.

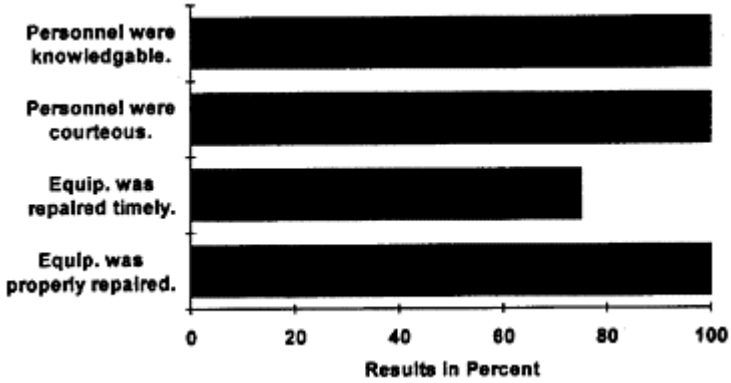


FIGURE 5.8 Customer satisfaction survey August-September 1994.

5.7 Teams

A *team* is a formal group of persons organized by the company to work together to accomplish certain goals and objectives. Normally, when teams are used in quality improvement programs, they are designed to achieve the organization's vision. The organization's vision is a statement of the desired end state of the organization articulated and deployed by the executive leadership. Organizational visions are inspiring, clear, challenging, reasonable, and empowering. Effective visions honor the past, while they prepare for the future. The following are types of teams that are being used in health care facilities today. Some of the names may not be common, but their definitions are very similar if not commensurate.

Process Action Teams (PAT)

Process action teams are composed of those who are involved in the process being investigated. The members of a PAT are often chosen by their respective managers. The primary consideration for PAT membership is knowledge about the operations of the organization and consequently the process being studied. The main function of a PAT is the performance of an improvement project. Hence customers are often invited to participate on the team. PATs use basic statistical and other tools to analyze a process and identify potential areas for improvement. PATs report their findings to an Executive Steering Committee or some other type of quality management improving group. [*"A problem well defined is half solved."* John Dewey, American philosopher and educator; 1859–1952.]

Transition Management Team (TMT)

The transition management team (see *Harvard Business Review*, Nov–Dec 1993, pp 109–118) is normally used for a major organizational change such as restructuring or reengineering. The TMT can be initiated due to the findings of a PAT, where it has been indicated that the process is severely broken and unsalvageable. The TMT is not a new layer of bureaucracy or a job for fading executives. The TMT oversees a large-scale corporate change effort. It makes sure that all change initiatives fit together. It is made up of 8 to 12 highly talented leaders who commit *all* their time making the transition a reality. The team members and what they are trying to accomplish must be accepted by the power structure of the organization. For the duration of the change process, they are the CEO's version of the National Guard. The CEO should be able to say, "I can sleep well tonight, the TMT is managing this." In setting up a TMT, organizations should adopt a fail-safe approach: Create a position to oversee the emotional and behavioral issues unless you can prove with confidence that you do not need one.

Quality Improvement Project Team (QIPT)

A quality improvement project team can be initiated due to the findings of a PAT, where it has been indicated that the process is broken. The main agenda of the QIPT is to improve the work process that managers have identified as important to change. The team studies this process methodically to find permanent solutions to problems. To do this, members can use many of the tools described in this chapter and in many other publications on quality and quality improvement available from schools, bookstores, and private organizations.

Executive Steering Committee (ESC)

This is an executive-level team composed of the Chief Executive Officer (CEO) of the organization and the executive staff that reports directly to the CEO. Whereas an organization may have numerous QMBs, PATs, and QIPTs, it has only one ESC. The ESC identifies strategic goals for organizational quality improvement efforts. It obtains information from customers to identify major product and service requirements. It is through the identification of these major requirements that quality goals for the organization are defined. Using this information, the ESC lists, prioritizes, and determines how to measure the organization's goals for quality improvement. The ESC develops the organization's improvement plan and manages the execution of that plan to ensure that improvement goals are achieved.

Quality Management Board (QMB)

This is a permanent cross-functional team made up of top and midlevel managers who are jointly responsible for a specific product, service or process. The structure of the board intended to improve communication and cooperation by providing vertical and horizontal "links" throughout the organization.

5.8 Process Improvement Model

This process is following the Joint Commission on the Accreditation of Healthcare Organizations' Quality Cube, a method of assessing the quality of the organization.

PLAN:

a. Identify the process to be monitored.
b. Select important functions and dimensions of performance applicable to the process identified.
c. Design a tool for collection of data.

MEASURE (Under this heading, you will document how, when, and where data was collected):

a. Collect data
b. Select the appropriate tool to deliver your data (charts, graphs, tables, etc.)

ASSESS (Document findings under this heading):

a. Interpret data collected.
b. Design and implement change.

(1) Redesign the process or tool if necessary.
(2) If no changes are necessary, then you have successfully used the Process Improvement pathway.

IMPROVEMENT (Document details here):

a. Set in place the process to gain and continue the improvement.

OUTCOME (Document all changes here):

a. Positive changes made to improve quality of care based on Performance Improvement Activity.

Problem-Solving Model

The FOCUS-PDCA/PMAIO Process Improvement Model is a statistics-based, quality-control method for improving processes. This approach to problem-solving could be used by all Process Action Teams to ensure uniformity within an organization. FOCUS-PDCA/PMAIO is as follows:

F—Find a process to improve.
O—Organize a team that knows the process.
C—Clarify current knowledge of the process.
U—Understand the cause or variations.
S—Select the process to improve
P—an the improvement. P—Plan
D—Do the improvement (pilot test). M—Measure
C—Check the results of the improvement. A—Assess
A—Act to hold the gain. I—Improve

O—Outcome

5.9 Summary

Although quality can be simply defined as conformity to customer or user requirements, it has many dimensions. Seven of them are described here: (1) performance, (2) aesthetics, (3) reliability (how dependably it performs), (4) availability (there when you need it), (5) durability (how long it lasts), (6) extras or features (supplementary items), and (7) serviceability (how easy it is to get serviced). The word *PARADES* can help you remember this:

PARADES: Seven Dimensions of Quality

1. *Performance:* A product or service that performs its intended function well scores high on this dimension of quality.
2. *Aesthetics:* A product or service that has a favorable appearance, sound, taste, or smell is perceived to be of good quality.
3. *Reliability:* Reliability, or dependability, is such an important part of product quality that quality-control engineers are sometimes referred to as *reliability engineers*.
4. *Availability:* A product or service that is there when you need it.
5. *Durability:* Durability can be defined as the amount of use one gets from a product before it no longer functions properly and replacement seems more feasible than constant repair.
6. *Extras:* Feature or characteristics about a product or service that supplements its basic functioning (i.e., remote control dialing on a television).
7. *Serviceability:* Speed, courtesy, competence, and ease of repair are all important quality factors.

Quality Has a Monetary Value!

Good quality often pays for itself, while poor quality is expensive in both measurable costs and hidden costs. The hidden costs include loss of goodwill, including loss of repeat business and badmouthing of the firm. High-quality goods and services often carry a higher selling price than do those of low quality. This information is evidenced by several reports in the *Wall Street Journal, Forbes Magazine, Money Magazine, Business Week Magazine,* etc. A good example is the turnaround of Japanese product sales using quality methodologies outlined in *The Deming Guide to Quality and Competitive Position,* by Howard S. and Shelly J.Gitlow. As Dr. Demings has stated, *quality improvement* must be *continuous*!

Quality is never an accident; it is always the result of intelligent energy.

John Ruskin, 1819–1900
English art critic and historian
Seven Lamps of Architecture

References

Bittel LR. 1985. *Every Supervisor Should Know,* 5th ed, pp 455–456. New York, Gregg Division/McGraw-Hill.

DuBrin AJ. 1994. *Essentials of Management,* 3d ed. Cleveland, South-Western Publishing Co.

Duck JD. 1993. Managing change—The art of balancing. *Harvard Business Review,* November–December 1993, pp 109–118.

Gitlow HS, Gitlow SJ. 1987. *The Deming Guide to Quality and Competitive Position.* Englewood Cliffs, NJ, Prentice-Hall.

Goal/QPC. 1988. *The Memory Jogger. A Pocket Guide of Tools for Continuous Improvement.* Massachusetts, Goal/QPC.

Ishikawa K. 1991. *Guide to Quality Control.* New York, Quality Resources (Asian Productivity Organization, Tokyo, Japan).

Joint Commission on Accreditation of Healthcare Organizations—*Comprehensive Accreditation Manual for Hospitals—Official Handbook.* Library of Congress Number 96–076721 1998, Oakbrook Terrace, Illinois 60181.

Juran JM. 1979. *Quality Control Handbook,* 3d ed. New York, McGraw-Hill.

Katzenbach JR, Smith DK. 1994. *The Wisdom of Teams.* New York, Harper Business . A Division of Harper Collins Publishers.

Mizuno S. 1988. *Management for Quality Improvement. The Seven New QC Tools.* Boston, Productivity Press.

Sholters PR. 1993. *The Team Handbook. How to Use Teams to Improve Quality.* Madison, Wise, Joiner Associates, Inc.

Walton M. 1986. *The Deming Management Method.* New York, Putnam Publishing Group.

6
A Standards Primer for Clinical Engineers

Alvin Wald
Columbia Univeisity

6.1 Introduction

The development, understanding, and use of standards is an important component of a clinical engineer's activities. Whether involved in industry, a health care facility, governmental affairs, or commercial enterprise, one way or another, the clinical engineer will find that standards are a significant aspect of professional activities. With the increasing emphasis on health care cost containment and efficiency, coupled with the continued emphasis on patient outcome, standards must be viewed both as a mechanism to reduce expenses and as another mechanism to provide quality patient care. In any case, standards must be addressed in their own right, in terms of technical, economic, and legal implications.

It is important for the clinical engineer to understand fully how standards are developed, how they are used, and most importantly, how they affect the entire spectrum of health-related matters. Standards exist that address systems (protection of the electrical power distribution system from faults), individuals (means to reduce potential electric shock hazards), and protection of the environment (disposal of deleterious waste substances).

From a larger perspective, standards have existed since biblical times. In the Book of Genesis [Chap. 6, ver. 14], Noah is given a construction standard by God, "Make thee an ark of gopher wood; rooms shalt thou make in the ark, and shalt pitch it within and without with pitch." Standards for weights and measures have played an important role in bringing together human societies through trade and commerce. The earliest record of a standard for length comes from ancient Egypt, in Dynasty IV (circa 3000 B.C.). This length was the royal cubit, 20.620 inches (52.379 cm), as used in construction of the Great Pyramid.

The importance of standards to society is illustrated in the Magna Carta, presented by the English barons to King John in 1215 on the field at Runnymede. Article 35 states:

> "There shall be standard measures of wine, beer, and corn—the London quarter—throughout the whole of our kingdom, and a standard width of dyed, russet and halberject cloth—two ells within the selvedges; and there shall be standard weights also."

0-8493-1813-0/03/$0.00+$1.50

The principles of this article appear in the English Tower system for weight and capacity, set in 1266 by the assize of Bread and Ale Act:

> "An English penny called a sterling, round and without any clipping, shall weigh thirty-two wheatcorns in the midst of the ear; and twenty ounces a pound: and eight pounds do make a gallon of wine, and eight gallons of wine do make a bushell, which is the eighth part of a quarter."

In the United States, a noteworthy use of standards occurred after the Boston fire of 1689. With the aim of rapid rebuilding of the city, the town fathers specified that all bricks used in construction were to be 9×4×4 inches. An example of standardization to promote uniformity in manufacturing practices was the contract for 10,000 muskets awarded to Eli Whitney by President Thomas Jefferson in 1800. The apocryphal story is that Eli Whitney (better known to generations of grammar school children for his invention of the cotton gin) assembled a large number of each musket part, had one of each part randomly selected, and then assembled a complete working musket. This method of production, the complete interchangeability of assembly parts, came to be known as the "armory method," replacing hand crafting, which at that time had been the prevailing method of manufacturing throughout the world.

6.2 Definitions

A most general definition of a standard is given by Rowe [1983].

"A standard is a multi-party agreement for establishing an arbitrary criterion for reference."

Each word used in the definition by Rowe corresponds to a specific characteristic that helps to define the concept of a standard.

Multi means more than one party, organization, group, government, agency, or individual.

Agreement means that the concerned parties have come to some mutually agreed upon understanding of the issues involved and of ways to resolve them. This understanding has been confirmed via some mechanism such as unanimity, consensus, ballot, or other means that has been specified.

Establishing defines the purpose of the agreement—to create the standard and carry forth its provisions.

Arbitrary emphasizes an understanding by the parties that there are no absolute criteria in creating the standard. Rather, the conditions and values chosen are based on the most appropriate knowledge and conditions available at the time the standard was established.

Criteria are those features and conditions that the parties to the agreement have chosen as the basis for the standard. Not all issues may be addressed, but only those deemed, for whatever reasons, suitable for inclusion.

A different type of definition of a standard is given in The United States Office of Management and Budget Circular A-119:

> " …a prescribed set of rules, conditions, or requirements concerned with the definition of terms; classification of components; delineation of procedures; specifications of materials, performance, design, or operations; or measurement of quality and quantity in describing materials, products, systems, services, or practices."

A code is a compilation of standards relating to a particular area of concern, i.e., a collection of standards. For example, local government health codes contain standards relating to providing of health care to members of the community. A regulation is an organization's way of specifying that some particular standard must be adhered to. Standards, codes, and regulations may or may not have legal implications, depending on whether the promulgating organization is governmental or private.

6.3 Standards for Clinical Engineering

There is a continually growing body of standards that affect health care facilities, and hence clinical engineering. The practitioner of health care technology must constantly search out, evaluate, and apply appropriate standards. The means to reconcile the conflicts of technology, cost considerations, the different jurisdictions involved, and the implementation of the various standards is not necessarily apparent. One technique that addresses these concerns and has proven to yield a consistent practical approach is a structured framework of the various levels of standards. This hierarchy of standards is a conceptual model that the clinical engineer can use to evaluate and apply to the various requirements that exist in the procurement and use of health care technology.

Standards have different purposes, depending on their particular applications. A hierarchy of standards can be used to delineate those conditions for which a particular standard applies. There are four basic categories, any one or all of which may be in simultaneous operation. (1) Local or proprietary standards (perhaps more properly called regulations) are developed to meet the internal needs of a particular organization. (2) Common interest standards serve to provide uniformity of product or service throughout an industry or profession. (3) Consensus standards are agreements among interested participants to address an area of mutual concern. (4) Regulatory standards are mandated by an authority having jurisdiction to define a particular aspect of concern. In addition, there are two categories of standards adherence: (1) voluntary standards, which carry no inherent power of enforcement, but provide a reference point of mutual understanding, and (2) mandatory standards, which are incumbent upon those to whom the standard is addressed, and enforceable by the authority having jurisdiction.

The hierarchy of standards model can aid the clinical engineer in the efficient and proper use of standards. More importantly, it can provide standards developers, users, and the authorities having jurisdiction in these matters with a structure by which standards can be effectively developed, recognized, and used to the mutual benefit of all.

6.4 A Hierarchy of Standards

Local, or proprietary standards, are developed for what might be called internal use. An organization that wishes to regulate and control certain of its own activities issues its own standards. Thus, the standard is local in the sense that it is applied in a specific venue, and it is proprietary in that it is the creation of a completely independent administration. For example, an organization may standardize on a single type of an electrocardiograph monitor. This standardization can refer to a specific brand or model, or to specific functional or operational features. In a more formal sense, a local standard may often be referred to as an institutional Policy and Procedure. The policy portion is the why of it; the procedure portion is the how. It must be kept in mind that standards of this type that are too restrictive will limit innovation and progress, in that they cannot readily adapt to novel conditions. On the other hand, good local standards contribute to lower costs, operational efficiency, and a sense of coherence within the organization.

Sometimes, local standards may originate from requirements of a higher level of regulation. For example, the Joint Commission for Accreditation of Healthcare Organizations [JCAHO] (formerly the Joint Commission for Hospital Accreditation (JCAH), a voluntary organization, (but an organization that hospitals belong to for various reasons, e.g., accreditation, reimbursement, approval of training programs) does not set standards for what or how equipment should be used. Rather, the JCAHO requires that each hospital set its own standards on how equipment is selected, used, and maintained. To monitor compliance with this requirement, the JCAHO inspects whether the hospital follows its own standards. In one sense, the most damaging evidence that can be adduced against an organization (or an individual) is that it (he) did not follow its (his) own standards.

Common interest standards are based on a need recognized by a group of interested parties, which will further their own interests, individually or collectively. Such standards are generally accepted by affected interests without being made mandatory by an authority; hence they are one type of voluntary standard. These standards are often developed by trade or professional organizations to promote uniformity in a product or process. This type of standard may have no inducement to adherence except for the benefits to the individual participants. For example, if you manufacture a kitchen cabinet that is not of standard size, it will not fit into the majority of kitchens and thus it will not sell. Uniformity of screw threads is another example of how a product can be manufactured and used by diverse parties, and yet be absolutely interchangeable. More recently, various information transfer standards allow the interchange of computer-based information among different types of instruments and computers.

Consensus standards are those that have been developed and accepted in accordance with certain well-defined criteria so as to assure that all points of view have been considered. Sometimes, the adjective “consensus” is used as a modifier for a “voluntary standard.” Used in this context, consensus implies that all interested parties have been consulted and have come to a general agreement on the provisions of the standard. The development of a consensus standard follows an almost ritualistic procedure to insure that fairness and due process are maintained. There are various independent voluntary and professional organizations that sponsor and develop standards on a consensus basis (see

below). Each such organization has its own particular rules and procedures to make sure that there is a true consensus in developing a standard.

In the medical products field, standards are sometimes difficult to implement because of the independent nature of manufacturers and their high level of competition. A somewhat successful standards story is the adoption of the DIN configuration for EGG lead-cable connection by the Association for the Advancement of Medical Instrumentation [AAMI]. The impetus for this standard was the accidental electrocution of several children brought about by use of the previous industry standard lead connection (a bare metal pin, as opposed to the new recessed socket). Most (but not all) manufacturers of EGG leads and cables now adhere to this standard. Agreement on this matter is in sharp contrast to the inability of the health care manufacturing industry to implement a standard for EGG cable connectors. Even though a standard was written, the physical configuration of the connector is not necessarily used by manufacturers in production, nor is it demanded by medical users in purchasing. Each manufacturer uses a different connector, leading to numerous problems in supply and incompatibility for users. This is an example of a voluntary standard, which for whatever reasons, is effectively ignored by all interested parties.

However, even though there have been some failures in standardization of product features, there has also been significant progress in generating performance and test standards for medical devices. A number of independent organizations sponsor development of standards for medical devices. For example, the American Society for Testing and Materials [ASTM] has developed, “Standard Specification for Minimum Performance and Safety Requirements for Components and Systems of Anesthesia Gas Machines (F1161–88).” Even though there is no statutory law that requires it, manufacturers no longer produce, and thus hospitals can no longer purchase, anesthesia machines without the built-in safety features specified in this standard. The Association for the Advancement of Medical Instrumentation has sponsored numerous standards that relate to performance of specific medical devices, such as defibrillators, electrosurgical instruments, and electronic sphygmomanometers. These standards are compiled in the AAMI publication, “Essential Standards for Biomedical Equipment Safety and Performance.” The National Fire Protection Association [NFPA] publishes “Standard for Health Care Facilities (NFPA 99),” which covers a wide range of safety issues relating to facilities. Included are sections that deal with electricity and electrical systems, central gas and vacuum supplies, and environmental conditions. Special areas such as anesthetizing locations, laboratories, and hyperbaric facilities are addressed separately.

Mandatory standards have the force of law or other authority having jurisdiction. Mandatory standards imply that some authority has made them obligatory. Mandatory standards can be written by the authority having jurisdiction, or they can be adapted from documents prepared by others as proprietary or consensus standards. The authority having jurisdiction can be a local hospital or even a department within the hospital, a professional society, a municipal or state government, or an agency of the federal government that has regulatory powers.

In the United States, hospitals are generally regulated by a local city or county authority, and/or by the state. These authorities set standards in the form of health codes or regulations, which have the force of law. Often, these local bodies consider the

requirements of a voluntary group, the Joint Commission for Accreditation of Healthcare Organizations, in their accreditation and regulatory processes.

American National Standards. The tradition in the United States is that of voluntary standards. However, once a standard is adopted by an organization, it can be taken one step further. The American National Standards Institute (ANSI) is a private, nongovernment, voluntary organization that acts as a coordinating body for standards development and recognition in the United States. If the development process for a standard meets the ANSI criteria of open deliberation of legitimate concerns, with all interested parties coming to a voluntary consensus, then the developers can apply (but are not required) to have their standard designated as an American National Standard. Such a designation does not make a standard any more legitimate, but it does offer some recognition as to the process by which it has been developed. ANSI also acts as a clearinghouse for standards development, so as to avoid duplication of effort by various groups that might be concerned with the same issues. ANSI is also involved as a U.S. coordinating body for many international standards activities.

An excellent source that lists existing standards and standards generating organizations, both nationally and internationally, along with some of the workings of the FDA (see below), is the *Medical Device Industry Fact Book* [Allen, 1996].

6.5 Medical Devices

On the national level, oversight is generally restricted to medical devices, and not on operational matters. Federal jurisdiction of medical devices falls under the purview of the Department of Health and Human Services, Public Health Service, Food and Drug Administration (FDA), Center for Devices and Radiological Health. Under federal law, medical devices are regulated under the "Medical Device Amendments of 1976" and the "Radiation Control for Health and Safety Act of 1968." Additional regulatory authorization is provided by the "Safe Medical Devices Act of 1990," the "Medical Device Amendments of 1992," the "FDA Reform and Enhancement Act of 1996," and the "Food and Drug Administration Modernization Act of 1997."

A medical device is defined by Section 201 of the Federal Food, Drug, and Cosmetic Act (as amended), as an:

> instrument, apparatus, implement, machine, contrivance, implant, *in vitro* reagent, or other similar or related article including any component, part, or accessory which is:
>
> recognized in the official National Formulary, or the United States Pharmacopeia, or any supplement to them;
>
> intended for use in the diagnosis of disease or other conditions, or in the care, mitigation, treatment, or prevention of disease, in man or other animals, or

intended to affect the structure of any function of the body of man or other animals; and which does not achieve its primary intended purposes through chemical action within or

on the body of man… and which is not dependent upon being metabolized for the achievement of its primary intended purposes.

The major thrust of the FDA has been in the oversight of the manufacture of medical devices, with specific requirements based on categories of perceived risks. The 1976 Act (Section 513) establishes three classes of medical devices intended for human use:

> Class I. General controls regulate devices for which controls other than performance standards or premarket approvals are sufficient to assure safety and effectiveness. Such controls include regulations that (1) prohibit adulterated or misbranded devices; (2) require domestic device manufacturers and initial distributors to register their establishments and list their devices; (3) grant FDA authority to ban certain devices; (4) provide for notification of risks and of repair, replacement, or refund; (5) restrict the sale, distribution, or use of certain devices; and (6) govern Good Manufacturing Practices, records, and reports, and inspections. These minimum requirements apply also to Class II and Class III devices.
>
> Class II. Performance Standards apply to devices for which general controls alone do not provide reasonable assurance of safety and efficacy, and for which existing information is sufficient to establish a performance standard that provides this assurance. Class II devices must comply not only with general controls, but also with an applicable standard developed under Section 514 of the Act. Until performance standards are developed by regulation, only general controls apply.
>
> Class III. Premarket Approval applies to devices for which general controls do not suffice or for which insufficient information is available to write a performance standard to provide reasonable assurance of safety and effectiveness. Also, devices that are used to support or sustain human life or to prevent impairment of human health, devices implanted in the body, and devices that present a potentially unreasonable risk of illness or injury. New Class III devices, those not "substantially equivalent" to a device on the market prior to enactment (May 28, 1976), must have approved Premarket Approval Applications (Section 510 k).

Exact specifications for General Controls and Good Manufacturing Practices (GMP) are defined in various FDA documents. Aspects of General Controls include yearly manufacturer registration, device listing, and premarket approval. General Controls are also used to regulate adulteration, misbranding and labeling, banned devices, and restricted devices. Good Manufacturing Practices include concerns of organization and personnel; buildings and equipment; controls for components, processes, packaging, and labeling; device holding, distribution, and installation; manufacturing records; product evaluation; complaint handling; and a quality assurance program. Design controls for GMP were introduced in 1996. They were motivated by the FDA's desire to harmonize its requirements with those of a proposed international standard (ISO 13485). Factors that need to be addressed include planning, input and output requirements, review, verification and validation, transfer to production, and change procedures, all contained in a history file for each device. Device tracking is typically required for Class III life-

sustaining and implant devices, as well as postmarket surveillance for products introduced starting in 1991.

Other categories of medical devices include combination devices, in which a device may incorporate drugs or biologicals. Combination devices are controlled via intercenter arrangements implemented by the FDA.

Transitional devices refer to devices that were regulated as drugs, prior to the enactment of the Medical Device Amendments Act of 1976. These devices were automatically placed into Class III, but may be transferred to Class I or II.

A custom device may be ordered by a physician for his or her own use or for a specific patient. These devices are not generally available, and cannot be labeled or advertised for commercial distribution.

An investigational device is one that is undergoing clinical trials prior to premarket clearance. If the device presents a significant risk to the patient, an Investigational Device Exemption must be approved by the FDA. Information must be provided regarding the device description and intended use, the origins of the device, the investigational protocol, and proof of oversight by an Institutional Review Board to insure informed patient consent. Special compassionate or emergency use for a nonapproved device or for a nonapproved use can be obtained from the FDA under special circumstances, such as when there is no other hope for the patient.

Adverse Events. The Safe Medical Devices Act of 1990 included a provision by which both users and manufacturers (and distributors) of medical devices are required to report adverse patient events that may be related to a medical device. Manufacturers must report to the FDA if a device (a) may have caused or contributed to a death or serious injury, or (b) malfunctioned in such a way as would be likely to cause or contribute to a death or serious injury if the malfunction were to reoccur. Device users are required to notify the device manufacturer of reportable incidents, and must also notify the FDA in case of a device-related death. In addition, the FDA established a voluntary program for reporting device problems that may not have caused an untoward patient event, but which may have the potential for such an occurrence under altered circumstances.

New devices. As part of the General Controls requirements, the FDA must be notified prior to marketing any new (or modifying an existing) device for patient use. This premarket notification, called the 510(k) process after the relevant section in the Medical Device Amendments Act, allows the FDA to review the device for safety and efficacy.

There are two broad categories that a device can fall into. A device that was marketed prior to May 28, 1976 (the date that the Medical Device Amendments became effective) can continue to be sold. Also, a product that is "substantially equivalent" to a preamendment device can likewise be marketed. However, the FDA may require a premarket approval application for any Class III device (see below). Thus, these preamendment devices and their equivalents are approved by "grandfathering." (Premarket notification to the FDA is still required to assure safety and efficacy). Of course, the question of substantial equivalency is open to an infinite number of interpretations. From the manufacturer's perspective, such a designation allows marketing the device without a much more laborious and expensive premarket approval process.

A new device that the FDA finds is not substantially equivalent to a premarket device is automatically placed into Class III. This category includes devices that provide

functions or work through principles not present in preamendment devices. Before marketing, this type of device requires a Premarket Approval Application by the manufacturer, followed by an extensive review by the FDA. (However, the FDA can reclassify such devices into Class I or II, obviating the need for premarket approval). The review includes scientific and clinical evaluation of the application by the FDA and by a Medical Advisory Committee (composed of outside consultants). In addition, the FDA looks at the manufacturing and control processes to assure that all appropriate regulatory requirements are being adhered to. Clinical (use of real patients) trials are often required for Class III devices in order to provide evidence of safety and efficacy. To carry out such trials, an Investigational Device Exemption must be issued by the FDA.

The Food and Drug Administration Modernization Act of 1997, which amends section 514 of the Food, Drug, and Cosmetic Act, has made significant changes in the above regulations. These changes greatly simplify and accelerate the entire regulatory process. For example, the law exempts from premarket notification Class I devices that are not intended for a use that is of substantial importance in preventing impairment of human health, or that do not present a potential unreasonable risk of illness or injury. Almost 600 Class I generic devices have been so classified by the agency. In addition, the FDA will specify those Class II devices for which a 510(k) submission will also not be required.

Several other regulatory changes have been introduced by the FDA to simplify and speed up the approval process. So-called “third party” experts will be allowed to conduct the initial review of all Class I and low-to-intermediate risk Class II devices. Previously, the FDA was authorized to create standards for medical devices. The new legislation allows the FDA to recognize and use all or parts of various appropriate domestic and internationally recognized consensus standards that address aspects of safety and/ or effectiveness relevant to medical devices.

6.6 International Standards

Most sovereign nations have their own internal agencies to establish and enforce standards. However, in our present world of international cooperation and trade, standards are tending toward uniformity across national boundaries. This internationalization of standards is especially true since formation of the European Common Market. The aim here is to harmonize the standards of individual nations by promulgating directives for medical devices that address “Essential Requirements” [Freeman, 1993] (see below). Standards in other areas of the world (Asia, Eastern Europe) are much more fragmented, with each country specifying regulations for its own manufactured and imported medical devices.

There are two major international standards generating organizations, both based in Europe, the International Electro technical Commission (IEC) and the International Organization for Standardization (ISO). Nations throughout the world participate in the activities of these organizations.

The International Electrotechnical Commission [IEC], founded in 1906, oversees, on an international level, all matters relating to standards for electrical and electronic items. Membership in the IEC is held by a National Committee for each nation. The United States National Committee (USNC) for IEC was founded in 1907, and since 1931 has

been affiliated with ANSI. USNC has as its members representatives from professional societies, trade associations, testing laboratories, government entities, other organizations, and individual experts. The USNC appoints a technical advisor and a technical advisory group for each IEC Committee and Subcommittee to help develop a unified United States position. These advisory groups are drawn from groups that are involved in the development of related U.S. national standards.

Standards are developed by Technical Committees (TC), Subcommittees (SC), and Working Groups (WG). IEC TC 62, "Electrical Equipment in Medical Practice," is of particular interest here. One of the basic standards of this Technical Committee is document 601–1, "Safety of Medical Electrical Equipment, Part 1: General Requirements for Safety," 2nd Edition (1988) and its Amendment 1 (1991), along with Document 601–1–1, "Safety Requirements for Medical Electrical Systems" (1992).

The International Organization for Standardization [ISO] oversees aspects of device standards other than those related to electro technology. This organization was formed in 1946 with a membership comprised of the national standards organizations of 26 countries. There are currently some 90 nations as members. The purpose of the ISO is to "facilitate international exchange of goods and services and to develop mutual cooperation in intellectual, scientific, technological, and economic ability." ISO addresses all aspects of standards except for electrical and electronic issues, which are the purview of the International Electrotechnical Commission. ANSI has been the official United States representative to ISO since its inception. For each Committee or Subcommittee of the ISO in which ANSI participates, a U.S. Technical Advisory Group (TAG) is formed. The administrator of the TAG is, typically, that same U.S. organization that is developing the parallel U.S. standard.

Technical Committees (TC) of the ISO concentrate on specific areas of interest. There are Technical Committees, Subcommittees, Working Groups and Study Groups. One of the member national standards organizations serves as the Secretariat for each of these technical bodies.

One standard of particular relevancy to manufacturers throughout the world is ISO 9000. This standard was specifically developed to assure a total quality management program that can be both universally recognized and applied to any manufacturing process. It does not address any particular product or process, but is concerned with structure and oversight of how processes are developed, implemented, monitored, and documented. An independent audit must be passed by any organization to obtain ISO 9000 registration. Many individual nations and manufacturers have adopted this standard and require that any product that they purchase be from a source that is ISO 9000 compliant.

The European Union was, in effect, created by the Single Europe Act, (EC-92) as a region "without internal frontiers in which the free movement of goods, persons, and capital is ensured." For various products and classes of products, the European Commission issues directives with regard to safety and other requirements, along with the means for assessing conformity to these directives. Products that comply with the appropriate directives can then carry the CE mark. EU member states ratify these directives into national law.

Two directives related to medical devices are the Medical Devices Directive (MDD), enacted in 1993 (mandatory as of June 15, 1998), and the Active Implanted Medical

Devices Directive (AIMDD), effective since 1995. Safety is the primary concern of this system, and as in the United States, there are three classes of risk. These risks are based on what and for how long the device touches, and its effects. Safety issues include electrical, mechanical, thermal, radiation, and labeling. Voluntary standards that address these issues are formulated by the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC).

6.7 Compliance with Standards

Standards that were originally developed on a voluntary basis may take on mandatory aspects. Standards that were developed to meet one particular need may be used to satisfy other needs as well. Standards will be enforced and adhered to if they meet the needs of those who are affected by them. For example, consider a standard for safety and performance for a defibrillator. For the manufacturer, acceptance and sales are a major consideration in both the domestic and international markets. People responsible for specifying, selecting, and purchasing equipment may insist on adherence to the standard so as to guarantee safety and performance. The user, physician or other health care professional, will expect the instrument to have certain operational and performance characteristics to meet medical needs. Hospital personnel want a certain minimum degree of equipment uniformity for ease of training and maintenance. The hospital's insurance company and risk manager want equipment that meets or exceeds recognized safety standards. Third party payers, that is private insurance companies or government agencies, insist on equipment that is safe, efficacious, and cost effective. Accreditation agencies, such as local health agencies or professional societies, often require equipment to meet certain standards. More basically, patients, workers, and society as a whole have an inherent right to fundamental safety. Finally, in our litigatious society, there is always the threat of civil action in the case of an untoward event in which a "nonstandard", albeit "safe", instrument was involved. Thus, even though no one has stated "this standard must be followed," it is highly unlikely that any person or organization will have the temerity to manufacture, specify, or buy an instrument that does not "meet the standard."

Another example of how standards become compulsory is via accreditation organizations. The Joint Commission for Accreditation of Healthcare Organizations has various standards (requirements). This organization is a private body that hospitals voluntarily accept as an accrediting agent. However, various health insurance organizations, governmental organizations, and physician specialty boards for resident education use accreditation by the JCAHO as a touchstone for quality of activities. Thus, an insurance company might not pay for care in a hospital that is not accredited, or a specialty board might not recognize resident training in such an institution. Thus, the requirements of the JCAHO, in effect, become mandatory standards for health care organizations.

A third means by which voluntary standards can become mandatory is by incorporation. Existing standards can be incorporated into a higher level of standards or codes. For example, various state and local governments incorporate standards developed by voluntary organizations, such as the National Fire Protection Association, into their own building and health codes. These standards then become, in effect, mandatory

government regulations, and have the force of (civil) law. In addition, as discussed above, the FDA will now recognize voluntary standards developed by recognized organizations.

6.8 Limitations of Standards

Standards are generated to meet the expectations of society. They are developed by organizations and individuals to meet a variety of specific needs, with the general goals of promoting safety and efficiency. However, as with all human activities, problems with the interpretation and use of standards do occur. Engineering judgment is often required to help provide answers. Thus, the clinical engineer must consider the limits of standards, a boundary that is not clear and is constantly shifting. Yet clinical engineers must always employ the highest levels of engineering principles and practices. Some of the limitations and questions of standards and their use will be discussed below.

Noncompliance with a Standard. Sooner or later, it is likely that a clinical engineer will either be directly involved with or become aware of deviation from an accepted standard. The violation may be trivial, with no noticeable effect, or there may be serious consequences. In the former case, either the whole incident may be ignored, or nothing more may be necessary than a report that is filed away, or the incident can trigger some sort of corrective action. In the latter case, there may be major repercussions involving investigation, censure, tort issues, or legal actions. In any event, lack of knowledge about the standard is not a convincing defense. Anyone who is in a position that requires knowledge about a standard should be fully cognizant of all aspects of that standard. In particular, one should know the provisions of the standard, how they are to be enforced, and the potential risks of noncompliance. Nonetheless, noncompliance with a standard, in whole or in part, may be necessary to prevent a greater risk or to increase a potential benefit to the patient. For example, when no other recourse is available, it would be defensible to use an electromagnet condemned for irreparable excessive leakage current to locate a foreign body in the eye of an injured person, and thus save the patient's vision. Even if the use of this device resulted in a physical injury or equipment damage, the potential benefit to the patient is a compelling argument for use of the noncompliant device. In such a case, one should be aware of and prepared to act on the possible hazard (excessive electrical current, here). A general disclaimer making allowance for emergency situations is often included in policy statements relating to use of a standard. Drastic conditions require drastic methods.

Standards and the Law. Standards mandated by a government body are not what is called "black-letter law", that is a law actually entered into a criminal or civil code. Standards are typically not adopted in the same manner as laws, i.e., they are not approved by a legislative body, ratified by an elected executive, and sanctioned by the courts. The usual course for a mandated standard is via a legislative body enacting a law that establishes or assigns to an executive agency the authority to regulate the concerned activities. This agency, under the control of the executive branch of government, then issues standards that follow the mandate of its enabling legislation. If conflicts arise, in addition to purely legal considerations, the judiciary must interpret the intent of the legislation in comparison with its execution. This type of law falls under civil rather than criminal application.

The penalty for noncompliance with a standard may not be criminal or even civil prosecution. Instead, there are administrative methods of enforcement, as well as more subtle yet powerful methods of coercion. The state has the power (and the duty) to regulate matters of public interest. Thus, the state can withhold or withdraw permits for construction, occupancy, or use. Possibly more effective, the state can withhold means of finance or payments to violators of its regulations. Individuals injured by failure to abide by a standard may sue for damages in civil proceedings. However, it must be recognized that criminal prosecution is possible when the violations are most egregious, leading to human injury or large financial losses.

Incorporation and Revision. Because of advances in technology and increases in societal expectations, standards are typically revised periodically. For example, the National Fire Protection Association revises and reissues its "Standard for Health Care Facilities" (NFPA 99) every 3 years. Other organizations follow a 5-year cycle of review, revision, and reissue of standards. These voluntary standards, developed in good faith, may be adapted by governmental agencies and made mandatory, as discussed above. When a standard is incorporated into a legislative code, it is generally referenced as to a particular version and date. It is not always the case that a newer version of the standard is more restrictive. For example, ever since 1984, the National Fire Protection Association "Standard for Health Care Facilities" (NFPA 99) does not require the installation of isolated power systems (isolation transformers and line isolation monitors) in anesthetizing locations that do not use flammable anesthetic agents, or in areas that are not classified as wet locations. A previous version of this standard, "Standard for the Use of Inhalation Anesthetics, (Flammable and Nonflammable)" (NFPA 56A-1978) did require isolated power. However, many State Hospital Codes have incorporated, by name and date, the provisions of the older standard, NFPA 56A. Thus, isolated power may still be required, by code, in new construction of all anesthetizing locations, despite the absence of this requirement in the latest version of the standard that addresses this issue. In such a case, the organization having jurisdiction in the matter must be petitioned to remedy this conflict between new and old versions of the standard.

Safety. The primary purpose of standards in clinical practice is to assure the safety of patient, operator, and bystanders. However, it must be fully appreciated that there is no such thing as absolute safety. The more safety features and regulations attached to a device, the less useful and the more cumbersome and costly may be its actual use. In the development, interpretation, and use of a standard, there are questions that must be asked: What is possible? What is acceptable? What is reasonable? Who will benefit? What is the cost? Who will pay?

No one can deny that medical devices should be made as safe as possible, but some risk will always remain. In our practical world, absolute safety is a myth. Many medical procedures involve risk to the patient. The prudent physician or medical technologist will recognize the possible dangers of the equipment and take appropriate measures to reduce the risk to a minimum. Some instruments and procedures are inherently more dangerous than others. The physician must make a judgment, based on his/her own professional knowledge and experience, as well as on the expectations of society, whether using a particular device is less of a risk than using an alternative device or doing nothing. Standards will help—but they do not guarantee complete safety, a cure, or legal and societal approval.

Liability. Individuals who serve on committees that develop standards, as well as organizations involved in such activities, are justifiably concerned with their legal position in the event that a lawsuit is instituted as a result of a standard that they helped to bring forth. Issues involved in such a suit may include restraint of trade, in case of commercial matters, or to liability for injury due to acts of commission or of omission. Organizations that sponsor standards or that appoint representatives to standards developing groups often have insurance for such activities. Independent standards committees and individual members of any standards committees may or may not be covered by insurance for participation in these activities. Although in recent times only one organization and no individual has been found liable for damages caused by improper use of standards (see following paragraph), even the possibility of being named in a lawsuit can intimidate even the most self-confident "expert." Thus, it is not at all unusual for an individual who is asked to serve on a standards development committee first to inquire as to liability insurance coverage. Organizations that develop standards or appoint representatives also take pains to insure that all of their procedures are carefully followed and documented so as to demonstrate fairness and prudence.

The dark side of standards is the implication that individuals or groups may unduly influence a standard to meet a personal objective, e.g., to dominate sales in a particular market. If standards are developed or interpreted unfairly, or if they give an unfair advantage to one segment, then restraint of trade charges can be made. This is why standards to be deemed consensus must be developed in a completely open and fair manner. Organizations that sponsor standards that violate this precept can be held responsible. In 1982, the United States Supreme Court, in the Hydrolevel Case [Perry, 1982], ruled that the American Society of Mechanical Engineers was guilty of antitrust activities because of the way some of its members, acting as a committee to interpret one of its standards, issued an opinion that limited competition in sales so as to unfairly benefit their own employers. This case remains a singular reminder that standards development and use must be inherently fair.

Inhibition. Another charge against standards is that they inhibit innovation and limit progress [Flink, 1984]. Ideally, standards should be written to satisfy minimum, yet sufficient, requirements for safety, performance, and efficacy. Improvements or innovations would still be permitted so long as the basic standard is followed. From a device user's point of view, a standard that is excessively restrictive may limit the scope of permissible professional activities. If it is necessary to abrogate a standard in order to accommodate a new idea or to extend an existing situation, then the choice is to try to have the standard changed, which may be very time consuming, or to act in violation of the standard and accept the accompanying risks and censure.

Ex Post Facto. A question continually arises as to what to do about old equipment (procedures, policies, facilities, etc.) when a new standard is issued or an old standard is revised so that existing items become obsolete. One approach, perhaps the simplest, is to do nothing, the philosophy here being that the old equipment was acquired in good faith and conformed to the then existing standards. As long as that equipment is usable and safe, there is no necessity to replace it. Another approach is to upgrade the existing equipment to meet the new standard. However, such modification may be technically impractical or financially prohibitive. Finally, one can simply throw out all of the existing equipment (or sell it to a second-hand dealer, or use the parts for maintenance) and buy

everything new. This approach would bring a smile of delight from the manufacturer and a scream of outrage from the hospital administrator. Usually what is done is a compromise, incorporating various aspects of these different approaches.

Costs. Standards cost both time and money to propose, develop, promulgate, and maintain. Perhaps the greatest hindrance to more participation in standards activities by interested individuals is the lack of funds to attend meetings where the issues are discussed and decisions are made. Unfortunately, but nonetheless true, organizations that can afford to sponsor individuals to attend such meetings have considerable influence in the development of that standard. On the other hand, those organizations that do have a vital interest in a standard should have an appropriate say in its development. A consensus of all interested parties tempers the undue influence of any single participant.

From another viewpoint, standards increase the costs of manufacturing devices, carrying out procedures, and administering policies. This incremental cost is, in turn, passed on to the purchaser of the goods or services. Whether or not the increased cost justifies the benefits of the standard is not always apparent. It is impossible to realistically quantify the costs of accidents that did not happen or the confusion that was avoided by adhering to a particular standard. However, it cannot be denied that standards have made a valuable contribution to progress, in the broadest sense of that word.

6.9 Conclusions

Standards are just like any other human activity; they can be well used or a burden. The danger of standards is that they will take on a life of their own, and rather than serve a genuine need will exist only as a justification of their own importance. This view is expressed in the provocative and iconoclastic book by Bruner and Leonard [1989], and in particular in their Chapter 9, “Codes and Standards: Who Makes the Rules?” However, the raison d’etre of standards is to do good. It is incumbent upon clinical engineers, not only to understand how to apply standards properly, but also how to introduce, modify, and retire standards as conditions change. Furthermore, the limitations of standards must be recognized in order to realize their maximum benefit. No standard can replace diligence, knowledge, and a genuine concern for doing the right thing.

References

Allen, A. (Ed.) 1996. *Medical Device Industry Fact Book* 3rd Edition. Canon Communications, Santa Monica, CA.

American Society for Testing and Materials [ASTM] 1916 Race Street, Philadelphia, PA 19103.

Association for the Advancement of Medical Instrumentation [AAMI] 3330 Washington Boulevard, Suite 400, Arlington, VA 22201.

Bruner, J.M.R. and Leonard, P.F. 1989. *Electricity, Safety and the Patient*, Year Book Medical Publishers, Chicago.

Flink, R. 1984. Standards: Resource or Constraint?, *IEEE Eng. Med. Biol. Mag.*, 3(1): 14–16.

Food and Drug Administration, Center for Devices and Radiological Health, 5600 Fishers Lane, Rockville, MD 20857. URL: http://www.fda.gov/

Freeman, M. 1993. The EC Medical Devices Directives, *IEEE Eng. Med. Biol. Mag.*, 12(2):79–80.

International Organization for Standardization [ISO], Central Secretariat, 1 rue de Varembe, Case postale 56, CH 1211, Geneva 20, Switzerland. URL: http://www.iso.ch/index.html

International Electrotechnical Commission [IEC], Central Office, 3 rue de Varembe, P.O. Box 131, CH-1211, Geneva 20, Switzerland. URL: http://www.iec.ch/

Joint Commission on Accreditation of Health Care Facilities [JCAHO] 1 Renaissance Boulevard, Oakbrook, IL 60181.

National Fire Protection Association [NFPA] Batterymarch Park, Quincy, MA 02269.

Perry, T.S. 1982. Antirust Ruling Chills Standards Setting, *IEEE Spectrum,* 19(8):52–54.

Rowe, W.D. 1983. Design and Performance Standards. In: *Medical Devices: Measurements, Quality Assurance, and Standards,* C.A.Caceres, H.T.Yolken, R.J.Jones, and H.R.Piehler, (Eds) p. 29–40, American Society for Testing and Materials, Philadelphia, PA.

7
Regulatory and Assessment Agencies

Mark E.Bruley
ECRI

Vivian H.Coates
ECRI

Effective management and development of clinical and biomedical engineering departments (herein-after called clinical engineering departments) in hospitals requires a basic knowledge of relevant regulatory and technology assessment agencies. Regulatory agencies set standards of performance and record keeping for the departments and the technology for which they are responsible. Technology assessment agencies are information resources for what should be an ever-expanding role of the clinical engineer in the technology decision-making processes of the hospital's administration.

This chapter presents an overview of regulatory and technology assessment agencies in the United States, Canada, Europe, and Australia that are germane to clinical engineering. Due to the extremely large number of such agencies and information resources, we have chosen to focus on those of greatest relevance and/or informational value. The reader is directed to the references and sources of further information presented at the end of the chapter.

7.1 Regulatory Agencies

Within the healthcare field, there are over 38,000 applicable standards, clinical practice guidelines, laws, and regulations [ECRI, 1999]. Voluntary standards are promulgated by more than 800 organizations; mandatory standards by more than 300 state and federal agencies. Many of these organizations and agencies issue guidelines that are relevant to the vast range of healthcare technologies within the responsibility of clinical engineering departments. Although many of these agencies also regulate the manufacture and clinical use of healthcare technology, such regulations are not directly germane to the management of a clinical department and are not presented.

For the clinical engineer, many agencies promulgate regulations and standards in the areas of, for example, electrical safety, fire safety, technology management, occupational safety, radiology and nuclear medicine, clinical laboratories, infection control, anesthesia and respiratory equipment, power distribution, and medical gas systems. In the U.S., medical device problem reporting is also regulated by many state agencies and by the U.S. Food and Drug Administration (FDA) via its MEDWATCH program. It is important

0-8493-1813-0/03/$0.00+$1.50

to note that, at present, the only direct regulatory authority that the FDA has over U.S. hospitals is in the reporting of medical device related accidents that result in serious injury or death. Chapter 6 discusses in detail many of the specific agency citations. Presented below are the names and addresses of the primary agencies whose codes, standards, and regulations have the most direct bearing on clinical engineering and technology management.

American Hospital Association
One North Franklin
Chicago, IL 60606
(312) 422–3000
Website: http://www.aha.org/

American College of Radiology
1891 Preston White Drive
Reston, VA 22091
(703) 648–8900
Website: http://www.acr.org/

American National Standards Institute
11 West 42nd Street
13th Floor
New York, NY 10036
(212) 642–4900
Website: http://www.ansi.org/

American Society for Hospital Engineering
840 North Lake Shore Drive
Chicago, IL 60611
(312) 280 5223
Website: http://www.ashe.org/

American Society for Testing and Materials
1916 Race Street
Philadelphia, PA 19103
(215) 299–5400
Website: http://www.astm.org/

Association for the Advancement of
Medical Instrumentation
3330 Washington Boulevard
Suite 400
Arlington, VA 22201
(703) 525–4890
Website: http://www.aami.org/

Australian Institute of Health and Welfare
GPO Box 570
Canberra, ACT 2601
Australia
(61) 06–243–5092
Website: http://www.aihw.gov.au/

British Standards Institution
2 Park Street
London, W1A 2BS
United Kingdom
(44) 071–629–9000
Website: http://www.bsi.org.uk/

Canadian Healthcare Association
17 York Street
Ottawa, ON KIN 9J6
Canada
(613) 241–8005
Website: http://www.canadian-healthcare.org/

CSA International
178 Rexdale Boulevard
Etobicoke, ON M9W 1R3
Canada
(416) 747–4000
Website: http://www.csa-international.org/

Center for Devices and Radiological Health
Food and Drug Administration
9200 Corporate Boulevard
Rockville, MD 20850
(301) 443–4690
Website: www.fda.gov/cdrh

Compressed Gas Association, Inc.
1725 Jefferson Davis Highway
Suite 1004
Arlington, VA 22202
(703) 412–0900

ECRI
5200 Butler Pike
Plymouth Meeting, PA 19462
(610) 825–6000
(610) 834–1275 fax

Websites: http://www.ecri.org/
http://www.ecriy2k.org/
http://www.mdsr.ecri.org/

Environmental Health Directorate
Health Protection Branch
Health Canada
Environmental Health Centre
19th Floor, Jeanne Mance Building
Tunney's Pasture
Ottawa, ON K1A 0L2 Canada
(613) 957–3143
Website: www.hc-sc.gc.ca/hpb/index_e.html

Therapeutic Products Programme
Health Canada
Holland Cross, Tower B
2nd Floor
1600 Scott Street
Address Locator #3102D1

Ottawa, ON K1A 1B6
(613) 954–0288
Website: www.hc-sc.gc.ca/hpb-dgps/therapeut

Food and Drug Administration
MEDWATCH, FDA Medical Products
Reporting Program
5600 Fishers Lane
Rockville, MD 20857–9787
(800) 332–1088
Website: www.fda.gov/cdrh/mdr.html

Institute of Electrical
and Electronics Engineers
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08850–1331
(732) 562–3800
Website: http://www.standards.ieee.org/

International Electrotechnical Commission
Box 131
3 rue de Varembe, CH 1211
Geneva 20
Switzerland

(41) 022–919–0211
Website: http://www.iec.ch/

International Organization for Standardization
1 rue de Varembe
Case postale 56, CH 1211
Geneva 20
Switzerland
(41) 022–749–0111
Website: http://www.iso.ch/

Joint Commission on Accreditation
of Healthcare Organizations
One Renaissance Boulevard
Oakbrook Terrace, IL 60181
(630) 792–5600
Website: http://www.jcaho.org/

Medical Devices Agency
Department of Health
Room 1209
Hannibal House
Elephant and Castle
London, SE1 6TQ
United Kingdom
(44) 171–972–8143
Website: http://www.medical-devices.gov.uk/

National Council on Radiation
Protection and Measurements
7910 Woodmont Avenue, Suite 800
Bethesda, MD 20814
(310) 657–2652
Website: http://www.ncrp.com/

National Fire Protection Association
1 Batterymarch Park
PO Box 9101
Quincy, MA 02269–9101
(617) 770–3000
Website: http://www.nfpa.org/

Nuclear Regulatory Commission
11555 Rockville Pike
Rockville, MD 20852

(301) 492–7000
Website: http://www.nrc.gov/

Occupational Safety
and Health Administration
US Department of Labor
Office of Information
and Consumer Affairs
200 Constitution Avenue, NW
Room N3647
Washington, DC 20210
(202) 219–8151
Website: http://www.osha.gov/

ORKI
National Institute for Hospital and Medical
Engineering
Budapest dios arok 3, H-1125
Hungary
(33) 1–156–1522

Radiation Protection Branch
Environmental Health Directorate
Health Canada
775 Brookfield Road
Ottawa, ON K1A 1C1
Website: www.hc-sc.gc.ca/ehp/ehd/rpb

Russian Scientific and Research Institute
Russian Public Health Ministry
EKRAN, 3 Kasatkina Street
Moscow
Russia 129301
(44) 071–405–3474

Society of Nuclear Medicine, Inc.
1850 Samuel Morse Drive
Reston, VA 20190–5316
(703) 708–9000
Website: http://www.snm.org/

Standards Association of Australia
PO Box 1055
Strathfield, NSW 2135
Australia

(61) 02–9746–4700
Website: http://www.standards.org.au/

Therapeutic Goods Administration
PO Box 100
Wooden, ACT 2606
Australia
(61) 2–6232–8610
Website: www.health.gov.au/tga

Underwriters Laboratories, Inc.
333 Pfingsten Road
Northbrook, IL 60062–2096
(847) 272–8800
Website: http://www.ul.com/

VTT
Technical Research Center of Finland
Postbox 316
SF-33101 Tampere 10
Finland
(358) 31–163300
Website: http://www.vti.fi/

7.2 Technology Assessment Agencies

Technology assessment is the practical process of determining the value of a new or emerging technology in and of itself or against existing or competing technologies using safety, efficacy, effectiveness, outcome, risk management, strategic, financial, and competitive criteria. Technology assessment also considers ethics and law as well as health priorities and cost effectiveness compared with competing technologies. A "technology" is defined as devices, equipment, related software, drugs, biotechnologies, procedures, and therapies; and systems used to diagnose or treat patients. The processes of technology assessment are discussed in detail in Chapter 2.

Technology assessment is not the same as technology acquisition/procurement or technology planning. The latter two are processes for determining equipment vendors, soliciting bids, and systematically determining a hospital's technology-related needs based on strategic, financial, risk management, and clinical criteria. The informational needs differ greatly between technology assessment and the acquisition/procurement or planning processes. This section focuses on the resources applicable to technology assessment.

Worldwide, there are nearly 400 organizations (private, academic, and governmental), providing technology assessment information, databases, or consulting services. Some are strictly information clearinghouses, some perform technology assessment, and some do both. For those that perform assessments, the quality of the information generated

varies greatly from superficial studies to in-depth, well-referenced analytical reports. In 1997, the U.S. Agency for Health Care Policy and Research (AHCPR) designated 12 "Evidence-Based Practice Centers" (EPC) to undertake major technology assessment studies on a contract basis. Each of these EPCs are noted in the list below and general descriptions of each center may be viewed on the internet at the AHCPR Website http://www.ahcpr.gov/clinic/epc/.

Language limitations are a significant issue. In the ultimate analysis, the ability to undertake technology assessment requires assimilating vast amounts of information, most of which exists only in the English language. Technology assessment studies published by the International Society for Technology Assessment in Health Care (ISTAHC), by the World Health Organization, and other umbrella organizations are generally in English. The new International Health Technology Assessment database being developed by ECRI in conjunction with the U.S. National Library of Medicine contains more than 30,000 citations to technology assessments and related documents.

Below are the names, mailing addresses, and Internet Website addresses of some of the most prominent organizations undertaking technology assessment studies:

Agence Nationale pour le Develeppement
de l'Evaluation Medicale
159 Rue Nationale
Paris 75013
France
(33) 42–16–7272
Website: www.upml.fr/andem/andem.htm

Agencia de Evaluacion de Technologias Sanitarias
Ministerio de Sanidad y Consumo
Institute de Salud Carlos III, AETS
Sinesio Delgado 6, 28029 Madrid
Spain
(34) 1–323–4359
Website: www.isciii.es/aets

Alberta Heritage Foundation for
Medical Research
125 Manulife Place
10180–101 Street
Edmonton, AB T5J 345
(403) 423–5727
Website: http://www.ahfmr.ab.ca/

American Association of Preferred
Provider Organizations
601 13th Street, NW
Suite 370 South

Washington, DC 20005
(202) 347–7600

American Academy of Neurology
1080 Montreal Avenue
St. Paul, MN 55116–2791
(612) 695–2716
Website: http://www.aan.com/

American College of Obstetricians
and Gynecologists
409 12th Street, SW
Washington, DC 20024
(202) 863–2518
Website: http://www.acog.org/

Australian Institute of Health and Welfare
GPO Box 570
Canberra, ACT 2601
Australia
(61) 06–243–5092
Website: http://www.aihw.gov.au/

Battelle Medical Technology Assessment
and Policy Research Center (MEDTAP)
901 D Street, SW
Washington, DC 20024
(202) 479–0500
Website: http://www.battelle.org/

Blue Cross and Blue Shield Association
Technology Evaluation Center
225 N Michigan Avenue
Chicago, IL 60601–7680
(312) 297–5530
(312) 297–6080 (publications)
Website: www.bluecares.com/new/clinical
(An EPC of AHCPR)

British Columbia Office of Health
Technology Assessment
Centre for Health Services & Policy Research,
University of British Columbia
429–2194 Health Sciences Mall
Vancouver, BC V6T 1Z3
CANADA

(604) 822–7049
Website: http://www.chspr.ubc.ca/

British Institute of Radiology
36 Portland Place
London, WIN 4AT
United Kingdom
(44) 171–580–4085
Website: http://www.bir.org.uk/

Canadian Coordinating Office for Health Technology Assessment
110–955 Green Valley Crescent
Ottawa ON K2C 3V4
CANADA
(613) 226–2553
Website: http://www.ccohta.ca/

Canadian Healthcare Association
17 York Street
Ottawa, ON KIN 9J6
Canada
(613) 241–8005
Website: http://www.canadian-healthcare.org/

Catalan Agency for Health Technology Assessment
Travessera de les Corts 131–159
Pavello Avenue
Maria, 08028 Barcelona
Spain
(34) 93–227–29–00
Website: http://www.aatm.es/

Centre for Health Economics
University of York
York Y0l 5DD
United Kingdom
(44) 01904–433718
Website: http://www.york.ac.uk/

Center for Medical Technology Assessment
Linköping University
5183 Linköping, Box 1026 (551–11)
Sweden
(46) 13–281–000

Center for Practice and Technology Assessment
Agency for Health Care Policy
and Research (AHCPR)
6010 Executive Boulevard, Suite 300
Rockville, MD 20852
(301) 594–4015
Website: http://www.ahcpr.gov/

Committee for Evaluation and Diffusion
of Innovative Technologies
3 Avenue Victoria
Paris 75004
France
(33) 1–40–273–109

Conseil devaluation des technologies
de la sante du Quebec
201 Cremazie Boulevard East
Bur 1.01
Montreal, PQ H2M 1L2
Canada
(514) 873–2563
Website: http://www.msss.gouv.qc.ca/

Danish Hospital Institute
Landermaerket 10
Copenhagen K
Denmark DK1119
(45) 33–11–5777

Danish Medical Research Council
Bredgade 43
1260 Copenhagen
Denmark
(45) 33–92–9700

Danish National Board of Health
Amaliegade 13
PO Box 2020
Copenhagen K
Denmark DK1012
(45) 35–26–5400

Duke Center for Clinical Health Policy Research
Duke University Medical Center
2200 West Main Street, Suite 320

Durham, NC 27705
(919) 286–3399
Website: http://www.clinipol.mc.duke.edu/
(An EPC of AHCPR)

ECRI
5200 Butler Pike
Plymouth Meeting, PA 19462
(610) 825–6000
(610) 834–1275 fax
Websites: http://www.ecri.org/
http://www.ecriy2k.org/
http://www.mdsr.ecri.org/
(An EPC of AHCPR)

Finnish Office for Health Care Technology Assessment
PO Box 220
FIN-00531 Helsinki
Finland
(35) 89–3967–2296
Website: www.stakes.fi/finohta

Frost and Sullivan, Inc.
106 Fulton Street
New York, NY 10038–2786
(212) 233–1080
Website: http://www.frost.com/

Health Council of the Netherlands
PO Box 1236
2280 CE
Rijswijk
The Netherlands
(31) 70–340–7520

Health Services Directorate Strategies and Systems for Health
Health Promotion
Health Promotion and Programs Branch
Health Canada
1915B Tunney's Pasture
Ottawa, ON K1A 1B4
Canada
(613) 954–8629
Website: www.hc-sc.gc.ca/hppb/hpol

Health Technology Advisory Committee
121 East 7th Place, Suite 400
PO Box 64975
St. Paul, MN 55164–6358
(612) 282–6358

Hong Kong Institute of Engineers
9/F Island Centre
No. 1 Great George Street
Causeway Bay
Hong Kong

Institute for Clinical PET
7100-A Manchester Boulevard
Suite 300
Alexandria, VA 22310
(703) 924–6650
Website: http://www.icpet.org/

Institute for Clinical Systems Integration
8009 34th Avenue South
Minneapolis, MN 55425
(612) 883–7999
Website: http://www.icsi.org/

Institute for Health Policy Analysis
8401 Colesville Road, Suite 500
Silver Spring, MD 20910
(301) 565–4216

Institute of Medicine (U.S.)
National Academy of Sciences
2101 Constitution Avenue, NW
Washington, DC 20418
(202) 334–2352
Website: www.nas.edu/iom

International Network of Agencies for
Health Technology Assessment
c/o SBU, Box 16158
S-103 24 Stockholm
Sweden
(46) 08–611–1913
Website: www.sbu.se/sbu-site/links/inahta

Johns Hopkins Evidence-Based Practice Center
The Johns Hopkins Medical Institutions
2020 E Monument Street, Suite 2–600
Baltimore, MD 21205–2223
(410) 955–6953
Website: www.jhsph.edu/Departments/Epi/
(An EPC of AHCPR)

McMaster University Evidence-Based Practice Center
1200 Main Street West, Room 3H7
Hamilton, ON L8N 3Z5
Canada
(905) 525–9140 ext. 22520
Website: http://hiru.mcmaster.ca.epc/
(An EPC of AHCPR)

Medical Alley
1550 Utica Avenue, South
Suite 725
Minneapolis, MN 55416
(612) 542–3077
Website: http://www.medicalalley.org/

Medical Devices Agency
Department of Health
Room 1209
Hannibal House
Elephant and Castle
London, SE1 6TQ
United Kingdom
(44) 171–972–8143
Website: http://www.medical-devices.gov.uk/

Medical Technology Practice Patterns Institute
4733 Bethesda Avenue, Suite 510
Bethesda, MD 20814
(301) 652–4005
Website: http://www.mtppi.org/

MEDTAP International
7101 Wisconsin Avenue, Suite 600
Bethesda MD 20814
(301) 654–9729
Website: http://www.medtap.com/

MetaWorks, Inc.
470 Atlantic Avenue
Boston, MA 02210
(617) 368–3573 ext. 206
Website: http://www.metawork.com/
(An EPC of AHCPR)

National Institute of Nursing Research, NIH
31 Center Drive, Room 5B10. MSC 2178
Bethesda, MD 20892–2178
(301) 496–0207
Website: www.nih.gov/ninr

National Commission on Quality Assurance
2000 L Street NW, Suite 500
Washington. DC 20036
(202) 955–3500
Website: http://www.ncqa.org/

National Committee of Clinical Laboratory
Standards (NCCLS)
940 West Valley Road, Suite 1400
Wayne, PA 19087–1898
(610) 688–0100
Website: http://www.nccls.org/

National Coordinating Center for
Health Technology Assessment
Boldrewood (Mailpoint 728)
Univ of Southampton SO 16 7PX
United Kingdom
(44) 170–359–5642
Website: www.soton.ac.uk/~hta/address.htm

National Health and Medical Research Council
GPO Box 9848
Canberra, ACT
Australia
(61) 06–289–7019

New England Medical Center
Center for Clinical Evidence Synthesis
Division of Clinical Research
750 Washington Street, Box 63
Boston, MA 02111
(617) 636–5133

Website: www.nemc.org/medicine/ccr/cces.htm
(An EPC of AHCPR)

New York State Department of Health
Tower Building
Empire State Plaza
Albany, NY 12237
(518) 474–7354
Website: http://www.health.state.ny.us/

NHS Centre for Reviews and Dissemination
University of York
York Y0l 5DD
United Kingdom
(44) 01–904–433634
Website: http://www.york.ac.uk/

Office of Medical Applications of Research
NIH Consensus Program Information Service
PO Box 2577
Kensington, MD 20891
(301) 231–8083
Website: odp.od.nih.gov/consensus

Ontario Ministry of Health
Hepburn Block
80 Grosvenor Street
10th Floor
Toronto, ON M7A 2C4
(416) 327–4377

Oregon Health Sciences University
Division of Medical Informatics
and Outcomes Research
3181 SW Sam lackson Park Road
Portland, OR 97201–3098
(503) 494–4277
Website: www.ohsu.edu/epc
(An EPC of AHCPR)

Pan American Health Organization
525 23rd Street NW
Washington, DC 20037–2895
(202) 974–3222
Website: http://www.paho.org/

Physician Payment Review Commission (PPRC)
2120 L Street NW
Suite 510
Washington, DC 20037
(202) 653–7220

Prudential Insurance Company of America
Health Care Operations and Research
Division
56 N Livingston Avenue
Roseland, NJ 07068
(201) 716–3870

Research Triangle Institute
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC 27709–2194
(919) 541–6512
(919) 541–7480
Website: www.rti.org/epc/
(An EPC of AHCPR)

San Antonio Evidence-Based Practice Center
University of Texas Health Sciences Center
Department of Medicine
7703 Floyd Curl Drive
San Antonio, TX 78284–7879
(210) 617–5190
Website: www.uthscsa.edu/
(An EPC of AHCPR)

Saskatchewan Health
Acute and Emergency Services Branch
3475 Albert Street
Regina, SK S4S 6X6
(306) 787–3656

Scottish Health Purchasing
Information Centre
Summerfield House
2 Eday Road
Aberdeen AB15 6RE
Scotland
United Kingdom
(44) 0–1224–663–456, ext. 75246
Website: www.nahat.net/shpic

Servicio de Evaluacion de Technologias Sanitarias
Duque de Wellington 2
E01010 Vitoria-Gasteiz
Spain
(94) 518–9250
E-mail: osteba-san@ej-gv.es

Society of Critical Care Medicine
8101 E Kaiser Boulevard
Suite 300
Anaheim, CA 92808–2259
(714) 282–6000
Website: http://www.sccm.org/

Swedish Council on Technology Assessment
in Health Care
Box 16158
S-103 24 Stockholm
Sweden
(46) 08–611–1913
Website: http://www.sbu.se/

Southern California EPC-RAND
1700 Main Street
Santa Monica, CA 90401
(310) 393–0411 ext. 6669
Website: www.rand.org/organization/health/epc/
(An EPC of AHCPR)

Swiss Institute for Public Health
Health Technology Programme
Pfrundweg 14
CH-5001 Aarau
Switzerland
(41) 064–247–161

TNO Prevention and Health
PO Box 2215
2301 CE Leiden
The Netherlands
(31) 71–518–1818
Website: www.tno.nl/instit/pg/index.html

University HealthSystem Consortium
2001 Spring Road, Suite 700
Oak Brook, IL 60523

(630) 954–1700
Website: http://www.uhc.edu/

University of Leeds
School of Public Health
30 Hyde Terrace
Leeds L52 9LN
United Kingdom
Website: http://www.leeds.ac.uk/

USCF-Stanford University EPC
University of California, San Francisco
505 Parnassus Avenue, Room M-1490
Box 0132
San Francisco, CA 94143–0132
(415) 476–2564
Website: www.stanford.edu/group/epc/
(An EPC of AHCPR)

U.S. Office of Technology Assessment
(former address) 600 Pennsylvania Avenue SE
Washington, DC 20003
Note: OTA closed on 29 Sep, 1995. However, documents can be accessed via the internet at www.wws.princeton.edu/5ota/html2/cong.html. Also a complete set of OTA publications is available on CD-ROM; contact the U.S. Government Printing Office (http://www.gpo.gov/) for more information.

Veterans Administration
Technology Assessment Program
VA Medical Center (152M)
150 S Huntington Avenue
Building 4
Boston, MA 02130
(617) 278–4469
Website: www.va.gov/resdev

Voluntary Hospitals of America, Inc.
220 East Boulevard
Irving, TX 75014
(214) 830–0000

Wessex Institute of Health Research and Development
Boldrewood Medical School

Bassett Crescent East
Highfield, Southampton SO 16 7PX
United Kingdom
(44) 01–703–595–661
Website: www.soton.ac.uk/~wi/index.html

World Health Organization
Distribution Sales
CH 1211
Geneva 27
Switzerland 2476
(41) 22–791–2111
Website: http://www.who.ch/
Note: Publications are also available from the WHOPublications Center, USA, at (518) 435–9686

References

ECRI. *Healthcare Standards Official Directory.* ECRI, Plymouth Meeting, PA, 1999.

Eddy DM. *A Manual for Assessing Health Practices & Designing Practice Policies:* The explicit approach. American College of Physicians, Philadelphia, PA, 1992.

Goodman C, Ed. *Medical Technology Assessment Directory.* National Academy Press, Washington, D.C., 1988.

van Nimwegen Chr, Ed. *International List of Reports on Comparative Evaluations of Medical Devices.* TNO Centre for Medical Technology, Leiden, the Netherlands, 1993.

Marcaccio KY, Ed. *Gale Directory of Databases.* Volume 1: Online databases. Gale Research International, London, 1993.

Further Information

A comprehensive listing of healthcare standards and the issuing organizations is presented in the *Healthcare Standards Directory* published by ECRI. This directory is well organized by keywords, organizations and their standards, federal and state laws, legislation and regulations, and contains a complete index of names and addresses.

The *International Health Technology Assessment* database is produced by ECRI. A portion of the database is also available in the U.S. National Library of Medicine's new database called *HealthSTAR*. Internet access to *HealthSTAR* is through Website address http://igm.nlm.nih.gov/. A description of the database may be found at http://www.nlm.nih.gov/pubs/factsheets/healthstar.html.

8
Applications of Virtual Instruments in Health Care

Eric Rosow
Hartford Hospital

Joseph Adam
Premise Development Corporation

8.1 Applications of Virtual Instruments in Health Care

Virtual Instrumentation (which is defined in Chapter 32, "Virtual Instrumentation: Applications in Biomedical Engineering") allows organizations to effectively harness the power of the PC to access, analyze, and share information throughout the organization. With vast amounts of data available from increasingly sophisticated enterprise-level data sources, potentially useful information is often left hidden due to a lack of useful tools. Virtual instruments can employ a wide array of technologies such as multidimensional analyses and Statistical Process Control (SPC) tools to detect patterns, trends, causalities, and discontinuities to derive knowledge and make informed decisions.

Today's enterprises create vast amounts of raw data, and recent advances in storage technology, coupled with the desire to use this data competitively, has caused a data glut in many organizations. The healthcare industry in particular is one that generates a tremendous amount of data. Tools such as databases and spreadsheets certainly help manage and analyze this data; however databases, while ideal for extracting data are generally not suited for graphing and analysis. Spreadsheets, on the other hand, are ideal for analyzing and graphing data, but this can often be a cumbersome process when working with multiple data files. Virtual instruments empower the user to leverage the best of both worlds by creating a suite of user-defined applications that allow the end-user to convert vast amounts of data into information that is ultimately transformed into knowledge to enable better decision making.

This chapter will discuss several virtual instrument applications and tools that have been developed to meet the specific needs of healthcare organizations. Particular attention will be placed on the use of quality control and "performance indicators" that provide the ability to trend and forecast various metrics. The use of SPC within virtual instruments will also be demonstrated. Finally, a nontraditional application of virtual instrumentation will be presented in which a "peer review" application has been developed to allow members of an organization to actively participate in the Employee Performance Review process.

0-8493-1813-0/03/$0.00+$1.50

Example Application #1: The EndoTester™—A Virtual Instrument-Based Quality Control and Technology Assessment System for Surgical Video Systems

The use of endoscopic surgery is growing, in large part because it is generally safer and less expensive than conventional surgery, and patients tend to require less time in a hospital after endoscopic surgery. Industry experts conservatively estimate that about 4 million minimally invasive procedures were performed in 1996. As endoscopic surgery becomes more common, there is an increasing need to accurately evaluate the performance characteristics of endoscopes and their peripheral components.

The assessment of the optical performance of laparoscopes and video systems is often difficult in the clinical setting. The surgeon depends on a high-quality image to perform minimally invasive surgery, yet assurance of proper function of the equipment by biomedical engineering staff is not always straightforward. Many variables in both patient and equipment may result in a poor image. Equipment variables, which may degrade image quality, include problems with the endoscope, either with optics or light transmission. The light cable is another source of uncertainty as a result of optical loss from damaged fibers. Malfunctions of the charge coupled device (CCD) video camera are yet another source of poor image quality. Cleanliness of the equipment, especially lens surfaces on the endoscope (both proximal and distal ends) are particularly common problems. Patient variables make the objective assessment of image quality more difficult. Large operative fields and bleeding at the operative site are just two examples of patient factors that may affect image quality.

The evaluation of new video endoscopic equipment is also difficult because of the lack of objective standards for performance. Purchasers of equipment are forced to make an essentially subjective decision about image quality. By employing virtual instrumentation, a collaborative team of biomedical engineers, software engineers, physicians, nurses, and technicians at Hartford Hospital (Hartford, CT) and Premise Development Corporation (Avon, CT) have developed an instrument, the EndoTester™, with integrated software to quantify the optical properties of both rigid and flexible fiberoptic endoscopes. This easy-to-use optical evaluation system allows objective measurement of endoscopic performance prior to equipment purchase and in routine clinical use as part of a program of prospective maintenance.

The EndoTester™ was designed and fabricated to perform a wide array of quantitative tests and measurements. Some of these tests include: (1) Relative Light Loss, (2) Reflective Symmetry, (3) Lighted (Good) Fibers, (4) Geometric Distortion, and (5) Modulation Transfer Function (MTF). Each series of tests is associated with a specific endoscope to allow for trending and easy comparison of successive measurements.

Specific information about each endoscope (i.e., manufacturer, diameter, length, tip angle, department/ unit, control number, and operator), the reason for the test (i.e., quality control, pre/post repair, etc.), and any problems associated with the scope are also documented through the electronic record. In addition, all the quantitative measurements from each test are automatically appended to the electronic record for life-cycle performance analysis.

Figures 8.1 and 8.2 illustrate how information about the fiberoptic bundle of an endoscope can be displayed and measured. This provides a record of the pattern of lighted optical fibers for the endoscope under test. The number of lighted pixels will

depend on the endoscope's dimensions, the distal end geometry, and the number of failed optical fibers. New fiber damage to an endoscope will be apparent by comparison of the lighted fiber pictures (and histogram profiles) from successive tests. Statistical data is also available to calculate the percentage of working fibers in a given endoscope.

In addition to the two-dimensional profile of lighted fibers, this pattern (and all other image patterns) can also be displayed in the form of a three-dimensional contour plot. This interactive graph may be viewed from a variety of viewpoints in that the user can vary the elevation, rotation, size, and perspective controls.

Figure 8.2 illustrates how test images for a specific scope can be profiled over time (i.e., days, months, years) to identify degrading performance. This profile is also useful to validate repair procedures by comparing test images before and after the repair.

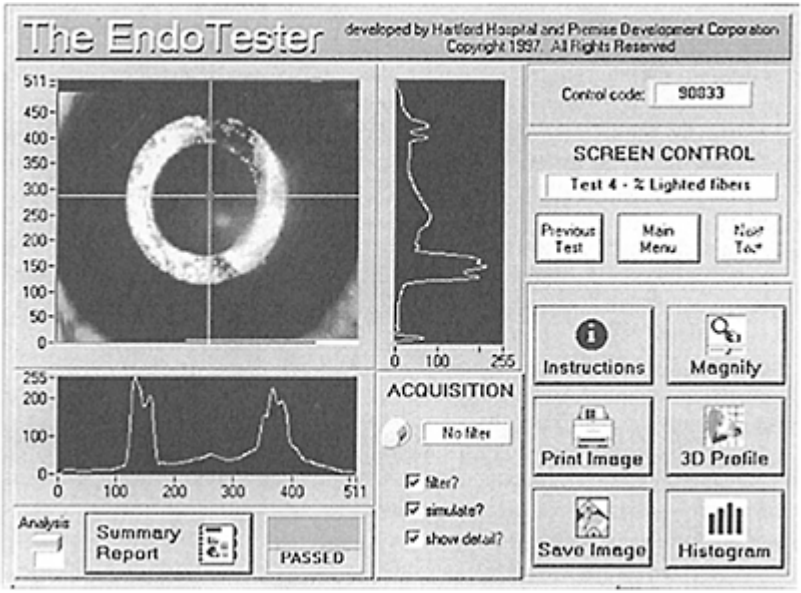


FIGURE 8.1 Endoscope tip reflection.

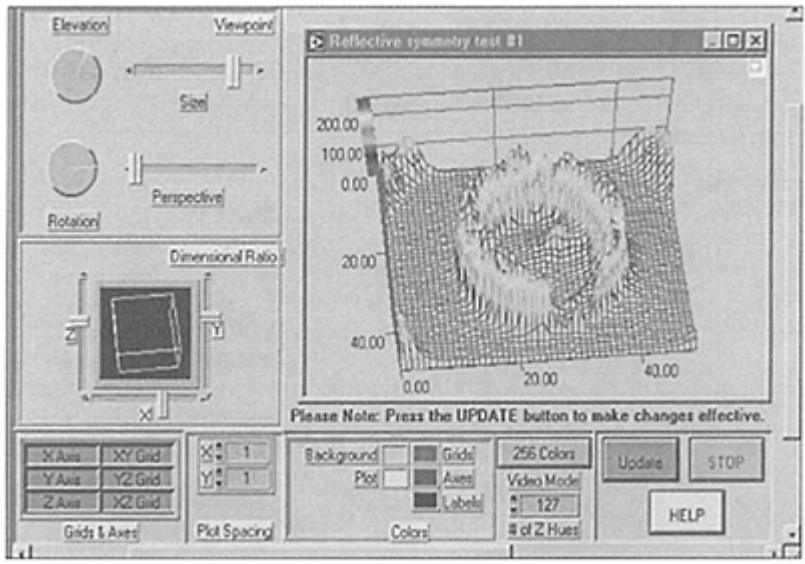


FIGURE 8.2 Endoscope profiling module.

The EndoTester™ has many applications. In general, the most useful application is the ability to objectively measure an endoscope's performance prior to purchase, and in routine clinical use as part of a program of prospective maintenance. Measuring parameters of scope performance can facilitate equipment purchase. Vendor claims of instrument capabilities can be validated as a part of the negotiation process. Commercially available evaluation systems (for original equipment manufacturers) can cost upward of $50,000, yet by employing the benefits of virtual instrumentation and a standard PC, an affordable, yet highly accurate test system for rigid and flexible fiberoptic endoscopes can now be obtained by clinical institutions.

In addition to technology assessment applications, the adoption of disposable endoscopes raises another potential use for the EndoTester™. Disposable scopes are estimated to have a life of 20 to 30 procedures. However, there is no easy way to determine exactly when a scope should be "thrown away." The EndoTester™ could be used to define this end-point.

The greatest potential for this system is as part of a program of preventive maintenance. Currently, in most operating rooms, endoscopes are removed from service and sent for repair when they fail in clinical use. This causes operative delay with attendant risk to the patient and an increase in cost to the institution. The problem is difficult because an endoscope may be adequate in one procedure but fail in the next that is more exacting due to clinical variables such as large patient size or bleeding. Objective assessment of endoscope function with the EndoTester™ may eliminate some of these problems.

Equally as important, an endoscope evaluation system will also allow institutions to ensure value from providers of repair services. The need for repair can be better defined

and the adequacy of the repair verified when service is completed. This ability becomes especially important as the explosive growth of minimally invasive surgery has resulted in the creation of a significant market for endoscope repairs and service. Endoscope repair costs vary widely throughout the industry, with costs ranging from $500 to $1500 or more per repair. Inappropriate or incomplete repairs can result in extending surgical time by requiring the surgeon to “switch scopes” (in some cases several times) during a surgical procedure.

Given these applications, we believe that the EndoTester™ can play an important role in reducing unnecessary costs, while at the same time improving the quality of the endoscopic equipment and the outcome of its utilization. It is the sincere hope of the authors that this technology will help to provide accurate, affordable and easy-to-acquire data on endoscope performance characteristics that clearly are to the benefit of the healthcare provider, the ethical service providers, manufacturers of quality products, the payers, and, of course, the patient.

Example Application #2: PIVIT™—Performance Indicator Virtual Instrument Toolkit

Most of the information management examples presented in this chapter are part of an application suite called PIVIT™. PIVIT is an acronym for “Performance Indicator Virtual Instrument Toolkit” and is an easy-to-use data acquisition and analysis product. PIVIT was developed specifically in response to the wide array of information and analysis needs throughout the healthcare setting.

PIVIT applies virtual instrument technology to assess, analyze, and forecast clinical, operational, and financial performance indicators. Some examples include applications that profile institutional indicators (i.e., patient days, discharges, percent occupancy, ALOS, revenues, expenses, etc.), and departmental indicators (i.e., salary, nonsalary, total expenses, expense per equivalent discharge, DRGs, etc.). Other applications of PIVIT include 360° Peer Review, Customer Satisfaction Profiling, and Medical Equipment Risk Assessment.

PIVIT can access data from multiple data sources. Virtually any parameter can be easily accessed and displayed from standard spreadsheet and database applications (i.e., Microsoft Access, Excel, Sybase, Oracle, etc.) using Microsoft’s Open Database Connectivity (ODBC) technology. Furthermore, multiple parameters can be profiled and compared in real time with any other parameter via interactive polar plots and three-dimensional displays. In addition to real-time profiling, other analyses such as SPC can be employed to view large data sets in a graphical format. SPC has been applied successfully for decades to help companies reduce variability in manufacturing processes. These SPC tools range from Pareto

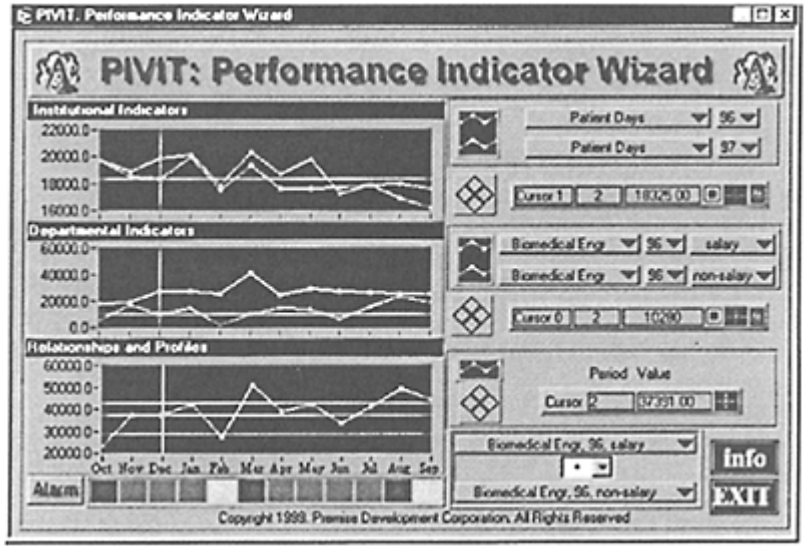


FIGURE 8.3 PIVIT™—Performance Indicator Wizard displays institutional and departmental indicators.

graphs to Run and Control charts. Although it will not be possible to describe all of these applications, several examples are provided below to illustrate the power of PIVIT.

Trending, Relationships, and Interactive Alarms

Figure 8.3 illustrates a virtual instrument that interactively accesses institutional and department specific indicators and profiles them for comparison. Data sets can be acquired directly from standard spreadsheet and database applications (i.e., Microsoft Access®, Excel®, Sybase®, Oracle®, etc.). This capability has proven to be quite valuable with respect to quickly accessing and viewing large sets of data. Typically, multiple data sets contained within a spreadsheet or database had to be selected and then a new chart of this data had to be created. Using PIVIT, the user simply selects the desired parameter from any one of the pull-down menus and this data set is instantly graphed and compared to any other data set.

Interactive “threshold cursors” dynamically highlight when a parameter is over and/or under a specific target. Displayed parameters can also be ratios of any measured value, for example, “Expense per Equivalent Discharge” or “Revenue to Expense Ratio.” The indicator color will change based on how far the data value exceeds the threshold value (i.e., from green to yellow to red). If multiple thresholds are exceeded, then the entire background of the screen (normally gray) will change to red to alert the user of an extreme condition.

Finally, multimedia has been employed by PIVIT to alert designated personnel with an audio message from the personal computer or by sending an automated message via e-mail, fax, pager, or mobile phone.

PIVIT also has the ability to profile historical trends and project future values. Forecasts can be based on user-defined history (i.e., “Months for Regression”), the type of regression (i.e., linear, exponential, or polynomial), the number of days, months, or years to forecast, and if any offset should be applied to the forecast. These features allow the user to create an unlimited number of “what-if” scenarios and allow only the desired range of data to be applied to a forecast. In addition to the graphical display of data values, historical and projected tables are also provided. These embedded tables look and function very much like a standard spreadsheet.

Data Modeling

Figure 8.4 illustrates another example of how virtual instrumentation can be applied to financial modeling and forecasting. This example graphically profiles the annual morbidity, mortality, and cost associated

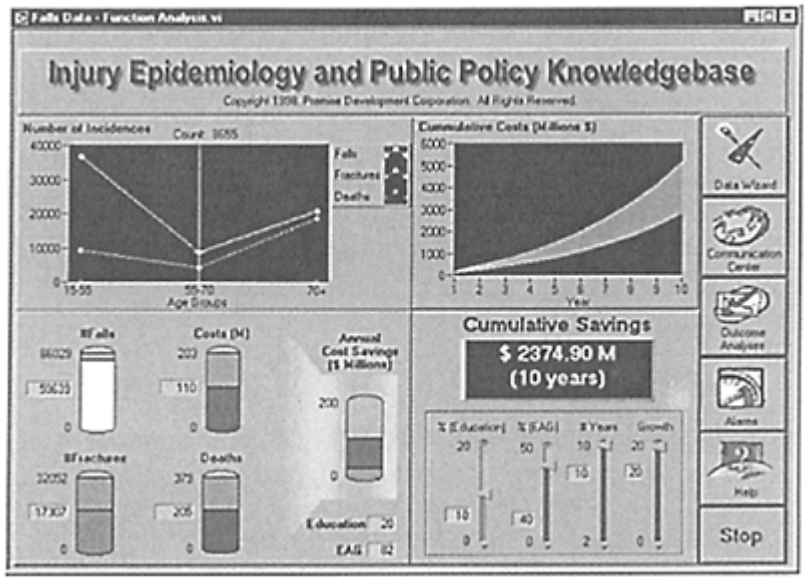


FIGURE 8.4 Injury epidemiology and public policy knowledgebase.

with falls within the state of Connecticut. Such an instrument has proved to be an extremely effective modeling tool due to its ability to interactively highlight relationships and assumptions, and to project the cost and/or savings of employing educational and other interventional programs.

Virtual instruments such as these are not only useful with respect to modeling and forecasting, but perhaps more importantly, they become a "knowledgebase" in which interventions and the efficacy of these interventions can be statistically proven. In addition, virtual instruments can employ standard technologies such as Dynamic Data Exchange (DDE), ActiveX, or TCP/IP to transfer data to commonly used software applications such as Microsoft Access® or Microsoft Excel®. In this way, virtual instruments can measure and graph multiple signals while at the same time send this data to another application that could reside on the network or across the Internet.

Another module of the PIVIT application is called the "Communications Center." This module can be used to simply create and print a report or it can be used to send e-mail, faxes, messages to a pager, or even leave voice-mail messages. This is a powerful feature in that information can be easily and efficiently distributed to both individuals and groups in real time.

Additionally, Microsoft Agent® technology can be used to pop up an animated help tool to communicate a message, indicate an alarm condition, or can be used to help the user solve a problem or point out a discrepancy that may have otherwise gone unnoticed. Agents employ a "text-to-speech" algorithm to actually "speak" an analysis or alarm directly to the user or recipient of the message. In this way, online help and user support can also be provided in multiple languages.

In addition to real-time profiling of various parameters, more advanced analyses such as SPC can be employed to view large data sets in a graphical format. SPC has been applied successfully for decades to help companies reduce variability in manufacturing processes. It is the opinion of this author that SPC has enormous applications throughout healthcare. For example, Fig. 8.5 shows how Pareto analysis can be applied to a sample trauma database of over 12,000 records. The Pareto chart may be frequency or percentage depending on front panel selection, and the user can select from a variety of different parameters by clicking on the "pull-down" menu. This menu can be configured to automatically display each database field directly from the database. In this example, various database fields (i.e., DRG, Principal Diagnosis, Town, Payer, etc.) can be selected for Pareto analysis. Other SPC tools include run charts, control charts, and process capability distributions.

Figure 8.6 illustrates a virtual instrument application that demonstrates how four "static" risk categories (and their corresponding values) are used to determine the inclusion of clinical equipment in the Medical Equipment Management Program at Hartford Hospital. Each risk category includes specific subcategories

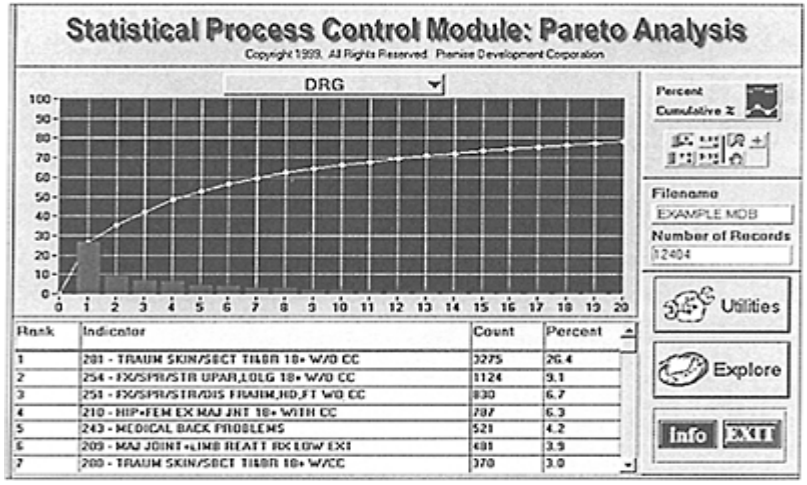


FIGURE 8.5 Statistical process control—Pareto analysis of a sample trauma registry.

Medical Equipment Risk Criteria

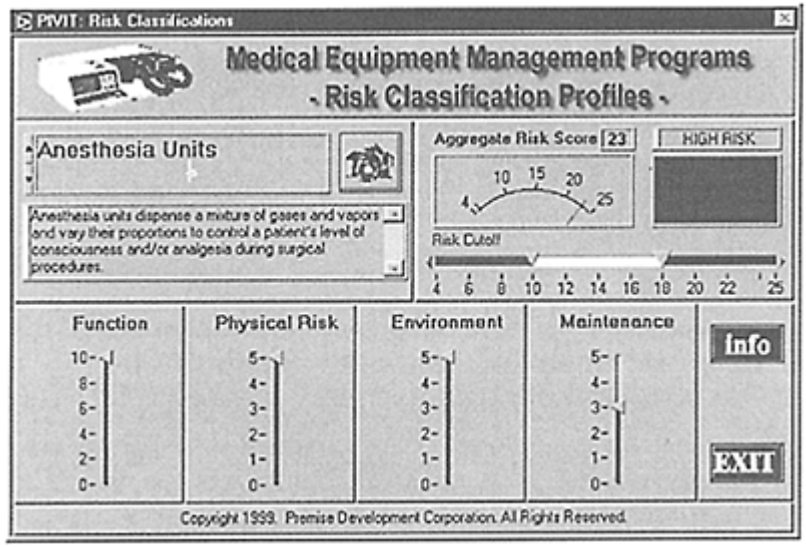


FIGURE 8.6 Medical equipment risk classification profiler.

that are assigned points, which when added together according to the formula listed below, yield a total score that ranges from 4 to 25.

Considering these scores, the equipment is categorized into five priority levels (High, Medium, Low, Grey List, and Noninclusion into the Medical Equipment Management Program). The four static risk categories are:

Equipment Function (EF): Stratifies the various functional categories (i.e., therapeutic, diagnostic, analytical, and miscellaneous) of equipment. This category has “point scores” that range from 1 (miscellaneous, nonpatient related devices) to 10 (therapeutic, life support devices).

Physical Risk (PR): Lists the “worst-case scenario” of physical risk potential to either the patient or the operator of the equipment. This category has “point scores” that range from 1 (no significant identified risk) to 5 (potential for patient and/or operator death).

Environmental Use Classification (EC): Lists the primary equipment area in which the equipment is used and has "point scores" that range from 1 (nonpatient care areas) to 5 (anesthetizing locations).

Preventive Maintenance Requirements (MR): Describes the level and frequency of required maintenance and has “point scores” that range from 1 (not required) to 5 (monthly maintenance).

The Aggregate Static Risk Score is calculated as follows:

Aggregate Risk Score=EF+PR+EC+MR

Using the criteria’s system described above, clinical equipment is categorized according to the following priority of testing and degree of risk:

High Risk: Equipment that scores between and including 18 to 25 points on the criteria’s evaluation system. This equipment is assigned the highest risk for testing, calibration, and repair.

Medium Risk: Equipment that scores between and including 15 to 17 points on the criteria’s evaluation system.

Low Risk: Equipment that scores between and including 12 to 14 points on the criteria’s evaluation system.

Hazard Surveillance (Gray): Equipment that scores between and including 6 and 11 points on the criteria’s evaluation system is visually inspected on an annual basis during the hospital hazard surveillance rounds.

Medical Equipment Management Program Deletion: Medical equipment and devices that pose little risk and scores less than 6 points may be deleted from the management program as well as the clinical equipment inventory.

Future versions of this application will also consider “dynamic” risk factors such as: user error, mean-time-between failure (MTBF), device failure within 30 days of a preventive maintenance or repair, and the number of years beyond the American Hospital Association's recommended useful life.

Peer Performance Reviews

The virtual instrument shown in Fig. 8.7 has been designed to easily acquire and compile performance information with respect to institution-wide competencies. It has been created to allow every member of a team or department to participate in the evaluation of

a co-worker (360° peer review). Upon running the application, the user is presented with a “Sign-In” screen where he or she enters their username and password. The application is divided into three components. The first (top section) profiles the employee and relevant service information. The second (middle section) indicates each competency as defined for employees, managers, and senior managers. The last (bottom) section allows the reviewer to evaluate performance by selecting one of four “radio buttons” and also provide specific comments related to each competency. This information is then compiled (with other reviewers) as real-time feedback.

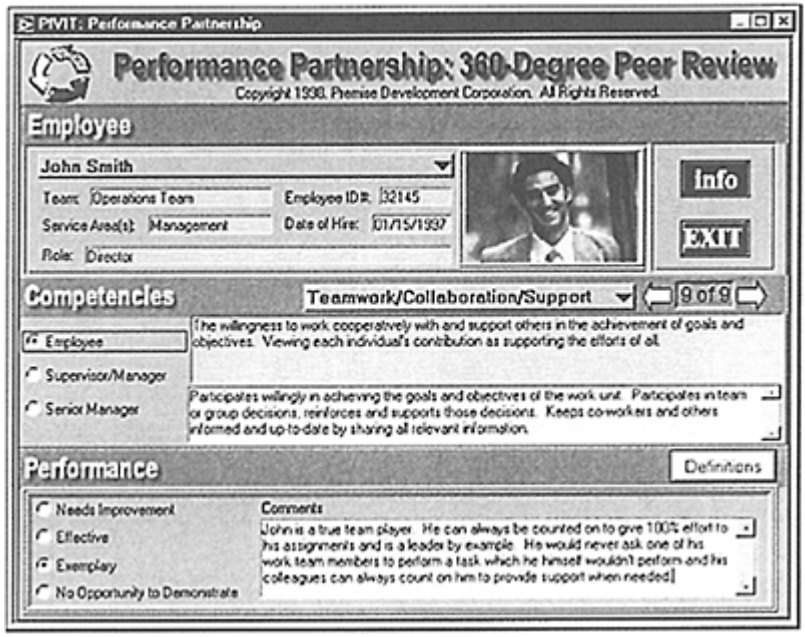


FIGURE 8.7 Performance reviews using virtual instrumentation.

References

1. American Society for Quality Control. American National Standard. *Definitions, Symbols, Forumulas, and Tables for Control Charts*, 1987.
2. Breyfogle, F W., *Statistical Methods for Testing, Development and Manufacturing*, John Wiley & Sons, New York, 1982.
3. Carey, R.G. and Lloyd, R.C. *Measuring Quality Improvement in Healthcare: A Guide to Statistical Process Control Applications*, 1995.
4. Fennigkow, L. and Lagerman, B., *Medical Equipment Management.* 1997 EC/PTSM Series/No. 1; Joint Commission on Accreditation of Hospital Organizations, 1997; 47–54.
5. *Frost & Sullivan Market Intelligence*, file 765, The Dialog Corporation, Worldwide Headquarters, 2440 W. El Camino Real, Mountain View, CA 94040.

6. Inglis, A., *Video Engineering*, McGraw Hill, New York, 1993.
7. Kutzner J., Hightower L., and Pruitt C. Measurement and testing of CCD sensors and cameras, *SMPTE Journal*, pp. 325–327, 1992.
8. Measurement of Resolution of Camera Systems, *IEEE Standard 208*, 1995.
9. *Medical Device Register 1997*, Volume 2, Montvale NJ, Medical Economics Data Production Company, 1997.
10. Montgomery, D.C., *Introduction to Statistical Quality Control*, 2nd ed., John Wiley & Sons, 1992.
11. Rosow, E., *Virtual Instrumentation Applications with BioSensors*, presented at the Biomedical Engineering Consortium for Connecticut (BEACON) Biosensor Symposium, Trinity College, Hartford, CT, October 2, 1998.
12. Rosow, E., Adam, J., and Beatrice, E, *The EndoTester™: A Virtual Instrument Endoscope Evaluation System for Fiberoptic Endoscopes*, *Biomedical Instrumentation and Technology*, pp. 480–487, September/October 1998.
13. Surgical Video Systems, *Health Devices*, 24(11), 428–457, 1995.
14. Walker, B., *Optical Engineering Fundamentals*, McGraw Hill, New York, 1995.
15. Wheeler, D.J. and Chambers, D.S., *Understanding Statistical Process Control*, SPG Press, 2nd ed., 1992.

II
Biomedical Sensors

Michael R.Neuman
Joint Program in Biomedical Engineering,
The University of Memphis and
University of Tennessee, Memphis

Historical Perspective: The Electrocardiograph **AP1-1**
Leslie A.Geddes The First Electrocardiogram • Capillary Electrometer Record • Rheotome Record • Mammalian Electrocardiograms • Corrected Capillary Electrometer Records • Clinical Electrocardiography

S SENSORS CONVERT SIGNALS OF ONE type of quantity such as hydrostatic fluid pressure into an equivalent signal of another type of quantity, for example, an electrical signal. Biomedical sensors take signals representing biomedical variables and convert them into what is usually an electrical signal. As such, the biomedical sensor serves as the interface between a biologic and an electronic system and must function in such a way as to not adversely affect either of these systems. In considering biomedical sensors, it is necessary to consider both sides of the interface: the biologic and the electronic, since both biologic and electronic factors play an important role in sensor performance.

TABLE II.1 Classifications of Biomedical Sensors

Physical sensors
Geometric
Mechanical
Thermal
Hydraulic
Electric
Optical
Chemical sensors
Gas
Electrochemical
Photometric
Other physical chemical methods
Bioanalytic

Many different types of Sensors Can be used in biomedical applications. Table II. 1 gives a general classification of these sensors. It is possible to cat egorize all sensors as being either physical or chemical. In the case of physical sensors, quantities such as geometric, mechanical, thermal, and hydraulic variables are measured. In biomedical applications these can include things such as mus cle displacement, blood pressure, core body ternperature, blood flow, cerebrospinal fluid pressure, and bone growth. Two types of physical sensors deserve special mention with regard to their biomedical application: Sensors of electrical phenom-ena in the body, usually known as electrodes, play a special

role as a result of their diagnostic and therapeutic applications. The most familiar of these are sensors used to pick up the electrocardiogram, an electrical signal produced by the heart. The other type of physical sensor that finds many applications in biology and medicine is the optical sensor. These sensors can use light to collect information, and, in the case of fiber-optic sensors, light is the signal transmission medium as well.

The second major classification of sensing devices is chemical sensors. In this case the sensors are concerned with measuring chemical quantities such as identifying the presence of particular chemical compounds, detecting the concentrations of various chemical species, and monitoring chemical activities in the body for diagnostic and therapeutic applications. A wide variety of chemical sensors can be classified in many ways. One such classification scheme is illustrated in Table II. 1 and is based upon the methods used to detect the chemical components being measured. Chemical composition can be measured in the gas phase using several techniques, and these methods are especially useful in biomedical measurements associated with the pulmonary system. Electrochemical sensors measure chemical concentrations or, more precisely, activities based on chemical reactions that interact with electrical systems. Photometric chemical sensors are optical devices that detect chemical concentrations based upon changes in light transmission, reflection, or color. The familiar litmus test is an example of an optical change that can be used to measure the acidity or alkalinity of a solution. Other types of physical chemical sensors such as the mass spectrometer use various physical methods to detect and quantify chemicals associated with biologic systems.

Although they are essentially chemical sensors, bioanalytic sensors are often classified as a separate major sensor category. These devices incorporate biologic recognition reactions such as enzyme-substrate, antigen-antibody, or ligand-receptor to identify complex biochemical molecules. The use of biologic reactions gives bioanalytic sensors high sensitivity and specificity in identifying and quantifying bio-chemical substances.

One can also look at biomedical sensors from the standpoint of their applications. These can be generally divided according to whether a sensor is used for diagnostic or therapeutic purposes in clinical medicine and for data collection in biomedical research. Sensors for clinical studies such as those carried out in the clinical chemistry laboratory must be standardized in such a way that errors that could result in an incorrect diagnosis or inappropriate therapy are kept to an absolute minimum. Thus these sensors must not only be reliable themselves, but appropriate methods must exist for testing the sensors that are a part of the routine use of the sensors for making biomedical measurements.

TABLE II.2 Types of Sensor-Subject Interfaces

Noncontacting (noninvasive)
Skin surface (contacting)
Indwelling (minimally invasive)
Implantable (invasive)

One can also look at biomedical sensors from the standpoint of how they are applied to the patient or research subject. Table II.2 shows the range of general approaches to attaching biomedical sensors. At the top of the have the method that involves the least

interaction with the biologic object being studied; the bottom of the list includes sensors that interact to the greatest extent. Clearly if a measurement can be made equally well by a sensor that does not contact the subject being measured or by one that must be surgically implanted, the former is by far the most desirable. However, a sensor that is used to provide information to help control a device surgically placed in the body to replace or assist a failing organ should be implanted, since this is the best way to communicate with the internal device.

You will notice in reading this section that the majority of biomedical sensors are essentially the same as sensors used in other applications. The unique part about biomedical sensors is their application. There are, however, special problems that are encountered by biomedical sensors that are unique to them. These problems relate to the interface between the sensor and the biologic system being measured. The presence of foreign materials, especially implanted materials, can affect the biologic environment in which they are located. Many biologic systems are designed to deal with foreign materials by making a major effort to eliminate them. The rejection reaction that is often discussed with regard to implanted materials or transplanted tissues is an example of this. Thus, in considering biomedical sensors, one must worry about this rejection phenomenon and how it will affect the performance of the sensor. If the rejection phenomenon changes the local biology around the sensor, this can result in the sensor measuring phenomena associated with the reaction that it has produced as opposed to phenomena characteristic of the biologic system being studied.

Biologic systems can also affect sensor performance. This is especially true for indwelling and implanted sensors. Biologic tissue represents a hostile environment that can degrade sensor performance. In addition to many corrosive ions, body fluids contain enzymes that break down complex molecules as a part of the body's effort to rid itself of foreign and toxic materials. These can attack the materials that make up the sensor and its package, causing the sensor to lose calibration or fail.

Sensor packaging is an especially important problem. The package must not only protect the sensor from the corrosive environment of the body, but it must allow that portion of the sensor that performs the actual measurement to communicate with the biologic system. Furthermore, because it is frequently desirable to have sensors be as small as possible, especially those that are implanted and indwelling, it is important that the packaging function be carried out without significantly increasing the size of the sensor structure. Although there have been many improvements in sensor packaging, this remains a major problem in biomedical sensor research. High-quality packaging materials that do not elicit major foreign body responses from the biologic system are still being sought.

Another problem that is associated with implanted sensors is that once they are implanted, access to them is very limited. This requires that these sensors be highly reliable so that there is no need to repair or replace them. It is also important that these sensors be highly stable, since in most applications it is not possible to calibrate the sensor *in vivo*. Thus, sensors must maintain their calibration once they are implanted, and for applications such as organ replacement, this can represent a potentially long time, the remainder of the patient's life.

In the following sections we will look at some of the sensors described above in more detail. We will consider physical sensors with special sections on biopotential electrodes

and optical sensors. We will also look at chemical sensors, including bioanalytic sensing systems. Although it is not possible to cover the field in extensive detail in a handbook such as this, it is hoped that these sections can serve as an introduction to this important aspect of biomedical engineering and instrumentation.

9
Physical Measurements

Michael R.Neuman
Joint Program in Biomedical Engineering, The University of Memphis and University of Tennessee, Memphis

Physical variables associated with biomedical systems are measured by a group of sensors known as physical sensors. Although many specific physical variables can be measured in biomedical systems, these can be categorized into a simple list as shown in Table 9.1. Sensors for these variables, whether they are measuring biomedical systems or other systems, are essentially the same. Thus, sensors of linear displacement can frequently be used equally well for measuring the displacement of the heart muscle during the cardiac cycle or the movement of a robot arm. There is, however, one notable exception regarding the similarity of these sensors: the packaging of the sensor and attachment to the system being measured. Although physical sensors used in nonbiomedical applications need to be packaged so as to be protected from their environment, few of these sensors have to deal with the harsh environment of biologic tissue, especially with the mechanisms inherent in this tissue for trying to eliminate the sensor as a foreign body. Another notable exception to this similarity of sensors for measuring physical quantities in biologic and nonbiologic systems are the sensors used for fluidic measurements such as pressure and flow. Special needs for these measurements in biologic systems have resulted in special sensors and instrumentation systems for these measurements that can be quite different from systems for measuing pressure and flow in nonbiologic environments.

In this chapter, we will attempt to review various examples of sensors used for physical measurement in biologic systems. Although it would be beyond the scope of this chapter to cover all these in detail, the principal sensors applied for biologic measurements will be described. Each section will include a brief description of the principle of operation of the sensor and the underlying physical principles, examples of some of the more common forms of these sensors for application in biologic systems, methods of signal processing for these sensors where appropriate, and important considerations for when the sensor is applied.

9.1 Description of Sensors

9.1.1 Linear and Angular Displacement Sensors

A comparison of various characteristics of displacement sensors described in detail below is outlined in Table 9.2.

0-8493-1813-0/03/$0.00+$1.50

Variable Resistance Sensor

One of the simplest sensors for measuring displacement is a variable resistor similar to the volume control on an audio electronic device [1]. The resistance between two terminals on this device is related to the

TABLE 9.1 Physical Variables and Sensors

Physical Quantity	Sensor	Variable Sensed
Geometric	Strain gauge	Strain
	LVDT	Displacement
	Ultrasonic transit time	Displacement
Kinematic	Velocimeter	Velocity
	Accelerometer	Acceleration
Force-Torque	Load cell	Applied force or torque
Fluidic	Pressure transducer	Pressure
	Flow meter	Flow
Thermal	Thermometer	Temperature
	Thermal flux sensor	Heat flux

TABLE 9.2 Comparison of Displacement Sensors

Sensor	Electrical Variable	Measurement Circuit	Sensitivity	Precision	Range
Variable resistor	Resistance	Voltage divider, ohmmeter, bridge, current source	High	Moderate	Large
Foil strain gauge	Resistance	Bridge	Low	Moderate	Small
Liquid metal strain gauge	Resistance	Ohmmeter, bridge	Moderate	Moderate	Large
Silicon strain gauge	Resistance	Bridge	High	Moderate	Small
Mutual inductance coils	Inductance	Impedance bridge, inductance meter	Moderate to high	Moderate to low	Moderate to large
Variable reluctance	Inductance	Impedance bridge, inductance meter	High	Moderate	Large
LVDT	Inductance	Voltmeter	High	High	High
Parallel plate capacitor	Capacitance	Impedance bridge, capacitance meter	Moderate to high	Moderate	Moderate to large
Sonic/ultrasonic	Time	Timer circuit	High	High	Large

linear or angular displacement of a sliding tap along a resistance element. Precision devices are available that have a reproducible, linear relationship between resistance and displacement. These devices can be connected in circuits that measure such resistance as an ohmmeter or bridge, or they can be used as a part of a circuit that provides a voltage that is proportional to the displacement. Such circuits include the voltage divider (as illustrated in Fig. 9.1*a*) or driving a known constant current through the resistance and measuring the resulting voltage across it. This sensor is simple and inexpensive and can be used for measuring relatively large displacements.

Strain Gauge

Another displacement sensor based on an electrical resistance change is the strain gauge [2]. If a long, narrow electrical conductor such as a piece of metal foil or a fine gauge wire is stretched within its elastic limit, it will increase in length and decrease in cross-sectional area. Because the electric resistance between both ends of this foil or wire can be given by

$$R = \rho \frac{l}{A} \tag{9.1}$$

where ρ is the electrical resistivity of the foil or wire material, l is its length, and A is its cross-sectional area, this stretching will result in an increase in resistance. The change in length can only be very small for the foil to remain within its elastic limit, so the change in electric resistance will also be small. The relative sensitivity of this device is given by its gauge factor, γ, which is defined as

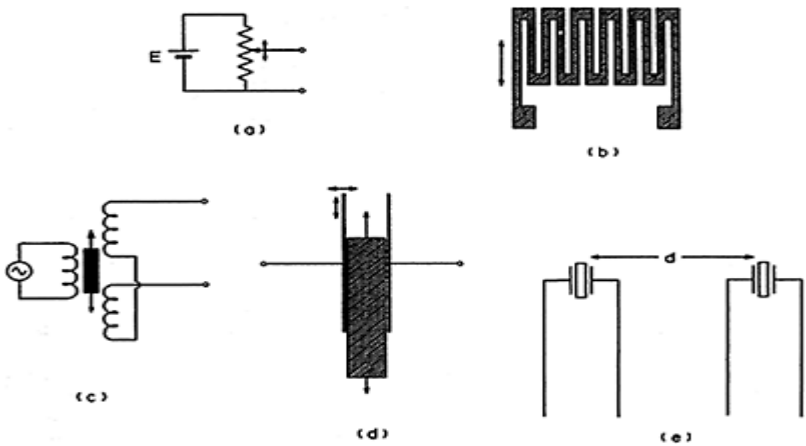


FIGURE 9.1 Examples of displacement sensors: *(a)* variable resistance sensor, *(b)* foil strain gauge, *(c)* linear variable differential transformer (LVDT), *(d)* parallel plate

capacitive sensor, and *(e)* ultrasonic transit time displacement sensor.

$$\gamma = \frac{\Delta R/R}{\Delta l/l}, \tag{9.2}$$

where ΔR is the change in resistance when the structure is stretched by an amount Δl. Foil strain gauges are the most frequently applied and consist of a structure such as shown in Fig. 9.1*b*. A piece of metal foil that is attached to an insulating polymeric film such as polyimide that has a much greater compliance than the foil itself is chemically etched into the pattern shown in Fig. 9.1*b*. When a strain is applied in the sensitive direction, the direction of the individual elements of the strain gauge, the length of the gauge will be slightly increased, and this will result in an increase in the electrical resistance seen between the terminals. Since the displacement or strain that this structure can measure is quite small for it to remain within its elastic limit, it can only be used to measure small displacements such as occur as loads are applied to structural beams. To increase the range of a foil strain gauge, it must be attached to some sort of a mechanical impedance converter such as a cantilever beam. If the strain gauge is attached to one surface of the beam as shown in Fig. 9.2, a fairly large displacement at the unsupported end of the beam can be translated to a relatively small displacement on the beam's surface. It would be possible for this structure to be used to measure larger displacements at the cantilever beam tip using a strain gauge on the beam.

Because the electric resistance changes for a strain gauge are quite small, the measurement of this resistance change can be challenging. Generally, Wheatstone bridge circuits are used. It is important to note, however, that changes in temperature can also result in electric resistance changes that are of the same order of magnitude or even larger than the electric resistance changes due to the strain. Thus, it is important to temperature-compensate strain gauges in most applications. A simple method of temperature compensation is to use a double or quadruple strain gauge and a bridge circuit for measuring the resistance change. This is illustrated in Fig. 9.2. When using the strain gauge in an application such as

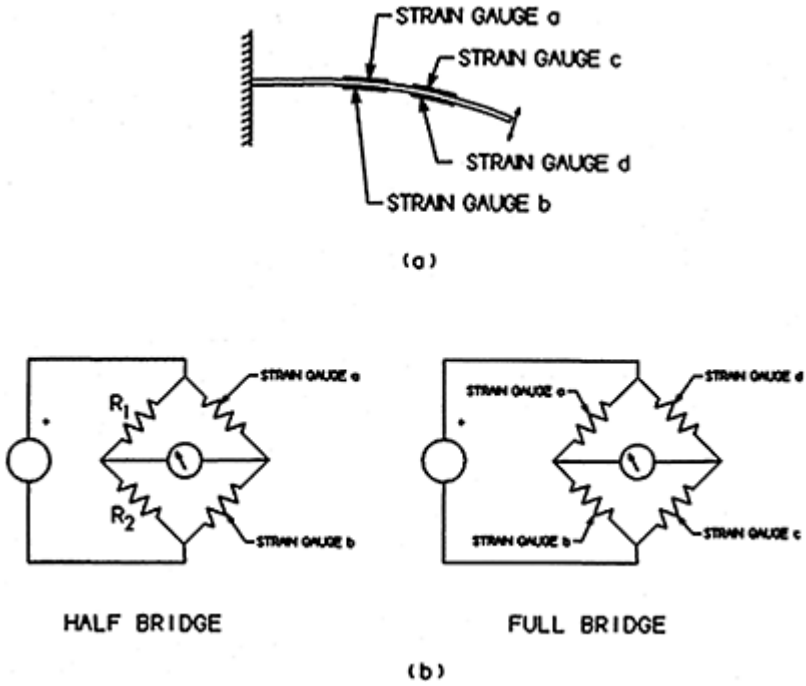


FIGURE 9.2 Strain gauges on a cantilever structure to provide temperature compensation: *(a)* cross-sectional view of the cantilever and *(b)* placement of the strain gauges in a half bridge or full bridge for temperature compensation and enhanced sensitivity.

the cantilever beam application described above, one or two of the strain gauge structures can be placed on the concave side of the beam and the other one or two on the convex side of the beam. Thus, as the beam deflects, the strain gauge on the convex side will experience tension, and that on the concave side will experience compression. By putting these gauges in adjacent arms of the Wheatstone bridge, their effects can double the sensitivity of the circuit in the case of the double strain gauge and quadruple it in the case where the entire bridge is made up of strain gauges on a cantilever.

In some applications it is not possible to place strain gauges so that one gauge is undergoing tension while the other is undergoing compression. In this case, the second strain gauge used for temperature compensation can be oriented such that its sensitive axis is in a direction where strain is minimal. Thus, it is still possible to have the temperature compensation by having two identical strain gauges at the same temperature

in adjacent arms of the bridge circuit, but the sensitivity improvement described in the previous paragraph is not seen.

Another constraint imposed by temperature is that the material to which the strain gauge is attached and the strain gauge both have temperature coefficients of expansion. Thus, even if a gauge is attached to a structure under conditions of no strain, if the temperature is changed, the strain gauge could experience some strain due to the different expansion that it will have compared to the structure to which it is attached. To avoid this problem, strain gauges have been developed that have identical temperature coefficients of expansion to various common materials. In selecting a strain gauge, one should choose a device with thermal expansion characteristics as close as possible to those of the object upon which the strain is to be measured.

A more compliant structure that has found applications in biomedical instrumentation is the liquid metal strain gauge [3]. Instead of using a solid electric conductor such as the wire or metal foil, mercury confined to a compliant, thin-wall, narrow-bore elastomer ic tube is used. The compliance of this strain gauge is determined by the elastic properties of the tube. Since only the elastic limit of the tube is of concern, this sensor can be used to detect much larger displacements than conventional strain gauges. Its sensitivity is roughly the same as a foil or wire strain gauge, but it is not as reliable. The mercury can easily become oxidized or small air gaps can occur in the mercury column. These effects make the sensor's characteristics noisy and sometimes results in complete failure.

Another variation on the strain gauge is the semiconductor strain gauge. These devices are frequently made out of pieces of silicon with strain gauge patterns formed using semiconductor micro-electronic technology. The principal advantage of these devices is that their gauge factors can be more than 50 times greater than that of the solid and liquid metal devices. They are available commercially, but they are a bit more difficult to handle and attach to structures being measured due to their small size and brittleness.

Inductance Sensors

Mutual Inductance

The mutual inductance between two coils is related to many geometric factors, one of which is the separation of the coils. Thus, one can create a very simple displacement sensor by having two coils that are coaxial but with different separation. By driving one coil with an ac signal and measuring the voltage signal induced in the second coil, this voltage will be related to how far apart the coils are from one another. When the coils are close together, the mutual inductance will be high, and so a higher voltage will be induced in the second coil; when the coils are more widely separated, the mutual inductance will be lower as will the induced voltage. The relationship between voltage and separation will be determined by the specific geometry of the coils and in general will not be a linear relationship with separation unless the change of displacement is relatively small. Nevertheless, this is a simple method of measuring separation that works reasonably well provided the coils remain coaxial. If there is movement of the coils transverse to their axes, it is difficult to separate the effects of transverse displacement from those of displacement along the axis.

Variable Reluctance

A variation on this sensor is the variable reluctance sensor wherein a single coil or two coils remain fixed on a form that allows a high reluctance slug to move into or out of the coil or coils along their axis. Since the position of this core material determines the number of flux linkages through the coil or coils, this can affect the self-inductance or mutual inductance of the coils. In the case of the mutual inductance, this can be measured using the technique described in the previous paragraph, whereas self-inductance changes can be measured using various instrumentation circuits used for measuring inductance. This method is also a simple method for measuring displacements, but the characteristics are generally nonlinear, and the sensor generally has only moderate precision.

Linear Variable Differential Transformer

By far the most frequently applied displacement transducer based upon inductance is the linear variable differential transformer (LVDT) [4]. This device is illustrated in Fig. 9.1*c* and is essentially a three-coil variable reluctance transducer. The two secondary coils are situated symmetrically about the primary coil and connected such that the induced voltages in each secondary oppose each other. When the core is located in the center of the structure equidistant from each secondary coil, the voltage induced in each secondary will be the same. Since these voltages oppose one another, the output voltage from the device will be zero. As the core is moved closer to one or the other secondary coils, the voltages in each coil will no longer be equal, and there will be an output voltage proportional to the displacement of the core from the central, zero-voltage position. Because of the symmetry of the structure, this voltage is linearly related to the core displacement. When the core passes through the central, zero point, the phase of the output voltage from the sensor changes by 180 degrees. Thus, by measuring the phase angle as well as the voltage, one can determine the position of the core. The circuit associated with the LVDT not only measures the voltage but often measures the phase angle as well.

LVDTs are available commercially in many sizes and shapes. Depending on the configuration of the coils, they can measure displacements ranging from tens of micrometers through centimeters.

Capacitive Sensors

Displacement sensors can be based upon measurements of capacitance as well as inductance. The fundamental principle of operation is the capacitance of a parallel plate capacitor as given by

$$C = \epsilon \frac{A}{d}, \tag{9.3}$$

where ϵ is the dielectric constant of the medium between the plates, d is the separation between the plates, and A is the cross-sectional area of the plates. Each of the quantities in Eq. (9.3) can be varied to form a displacement transducer as shown in Fig. 9.1*c*. By moving one of the plates with respect to the other, Eq. (9.3) shows us that the capacitance

will vary inversely with respect to the plate separation. This will give a hyperbolic capacitance-displacement characteristic. However, if the plate separation is maintained at a constant value and the plates are displaced laterally with respect to one another so that the area of overlap changes, this can produce a capacitance-displacement characteristic that can be linear, depending on the shape of the actual plates.

The third way that a variable capacitance transducer can measure displacement is by having a fixed parallel plate capacitor with a slab of dielectric material having a dielectric constant different from that of air that can slide between the plates. The effective dielectric constant for the capacitor will depend on how much of the slab is between the plates and how much of the region between the plates is occupied only by air. This, also, can yield a transducer with linear characteristics.

The electronic circuitry used with variable capacitance transducers is essentially the same as any other circuitry used to measure capacitance. As with the inductance transducers, this circuit can take the form of a bridge circuit or specific circuits that measure capacitive reactance.

Sonic and Ultrasonic Sensors

If the velocity of sound in a medium is constant, the time it takes a short burst of that sound energy to propagate from a source to a receiver will be proportional to the displacement between the two transducers. This is given by

$$d = cT\,, \tag{9.4}$$

where c is the velocity of sound in the medium, T is the transit time, and d is the displacement. A simple system for making such a measurement is shown in Fig. 9.1*e* [5]. A brief sonic or ultrasonic pulse is generated at the transmitting transducer and propagates through the medium. It is detected by the receiving transducer at time T after the burst was initiated. The displacement can then be determined by applying Eq. (9.4).

In practice, this method is best used with ultrasound, since the wavelength is shorter, and the device will neither produce annoying sounds nor respond to extraneous sounds in the environment. Small piezoelectric transducers to generate and receive ultrasonic pulses are readily available. The electronic circuit used with this instrument carries out three functions: (1) generation of the sonic or ultrasonic burst, (2) detection of the received burst, and (3) measurement of the time of propagation of the ultrasound. An advantage of this system is that the two transducers are coupled to one another only sonically. There is no physical connection as was the case for the other sensors described in this section.

Velocity Measurement

Velocity is the time derivative of displacement, and so all the displacement transducers mentioned above can be used to measure velocity if their signals are processed by passing them through a differentiator circuit. There are, however, two additional methods that can be applied to measure velocity directly.

Magnetic Induction

If a magnetic field that passes through a conducting coil varies with time, a voltage is induced in that coil that is proportional to the time-varying magnetic field. This relationship is given by

$$v = N\frac{d\phi}{dt} \tag{9.5}$$

where v is the voltage induced in the coil, N is the number of turns in the coil, and ϕ is the total magnetic flux passing through the coil (the product of the flux density and area within the coil). Thus a simple way to apply this principle is to attach a small permanent magnet to an object whose velocity is to be determined, and attach a coil to a nearby structure that will serve as the reference against which the velocity is to be measured. A voltage will be induced in the coil whenever the structure containing the permanent magnet moves, and this voltage will be related to the velocity of that movement. The exact relationship will be determined by the field distribution for the particular magnet and the orientation of the magnet with respect to the coil.

Doppler Ultrasound

When the receiver of a signal in the form of a wave such as electromagnetic radiation or sound is moving at a nonzero velocity with respect to the emitter of that wave, the frequency of the wave perceived by the receiver will be different than the frequency of the transmitter. This frequency difference, known as the Doppler shift, is determined by the relative velocity of the receiver with respect to the emitter and is given by

$$f_d = \frac{f_o u}{c} \tag{9.6}$$

where f_d is the Doppler frequency shift, f_o is the frequency of the transmitted wave, u is the relative velocity between the transmitter and receiver, and c is the velocity of sound in the medium. This principle can be applied in biomedical applications as a Doppler velocimeter. A piezoelectric transducer can be used as the ultrasound source with a similar transducer as the receiver. When there is no relative movement between the two transducers, the frequency of the signal at the receiver will be the same as that at the emitter, but when there is relative motion, the frequency at the receiver will be shifted according to Eq. (9.6).

The ultrasonic velocimeter can be applied in the same way that the ultrasonic displacement sensor is used. In this case the electronic circuit produces a continuous ultrasonic wave and, instead of detecting the transit time of the signal, now detects the frequency difference between the transmitted and received signals. This frequency difference can then be converted into a signal proportional to the relative velocity between the two transducers.

Accelerometers

Acceleration is the time derivative of velocity and the second derivative with respect to time of displacement. Thus, sensors of displacement and velocity can be used to determine acceleration when their signals are appropriately processed through differentiator circuits. In addition, there are direct sensors of acceleration based upon Newton's second law and Hooke's law. The fundamental structure of an accelerometer is shown in Fig. 9.3. A known seismic mass is attached to the housing by an elastic element. As the

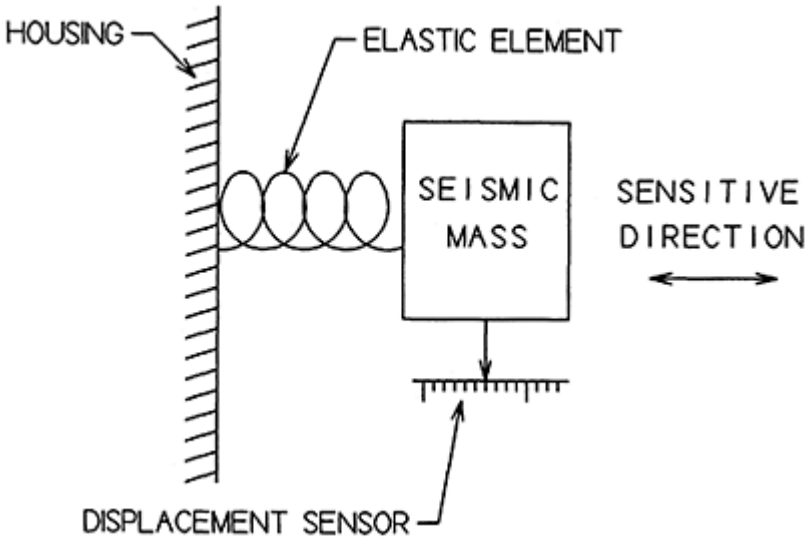


FIGURE 9.3 Fundamental structure of an accelerometer.

structure is accelerated in the sensitive direction of the elastic element, a force is applied to that element according to Newton's second law. This force causes the elastic element to be distorted according to Hooke's law, which results in a displacement of the mass with respect to the accelerometer housing. This displacement is measured by a displacement sensor. The relationship between the displacement and the acceleration is found by combining Newton's second law and Hooke's law

$$a = \frac{k}{m} x , \tag{9.7}$$

where x is the measured displacement, m is the known mass, k is the spring constant of the elastic element, and a is the acceleration. Any of the displacement sensors described above can be used in an accelerometer. The most frequently used displacement sensors are strain gauges or the LVDT. One type of accelerometer uses a piezoelectric sensor as both the displacement sensor and the elastic element. A piezoelectric sensor generates an

electric signal that is related to the dynamic change in shape of the piezoelectric material as the force is applied. Thus, piezoelectric materials can only directly measure time-varying forces. A piezoelectric accelerometer is, therefore, better for measuring changes in acceleration than for measuring constant accelerations. A principal advantage of piezoelectric accelerometers is that they can be made very small, which is useful in many biomedical applications.

Force

Force is measured by converting the force to a displacement and measuring the displacement with a displacement sensor. The conversion takes place as a result of the elastic properties of a material. Applying a force to the material distorts the material's shape, and this distortion can be determined by a displacement sensor. For example, the cantilever structure shown in Fig. 9.2*a* could be a force sensor. Applying a vertical force at the tip of the beam will cause the beam to deflect according to its elastic properties. This deflection can be detected using a displacement sensor such as a strain gauge as described previously.

A common form of force sensor is the load cell. This consists of a block of material with known elastic properties that has strain gauges attached to it. Applying a force to the load cell stresses the material, resulting in a strain that can be measured by the strain gauge. Applying Hooke's law, one finds that the strain is proportional to the applied force. The strain gauges on a load cell are usually in a half-bridge

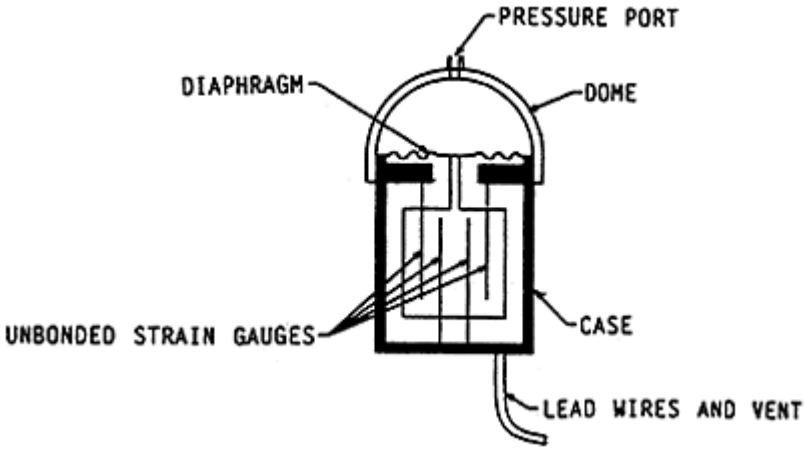


FIGURE 9.4 Structure of an unbonded strain gauge pressure sensor. Reproduced with permission from Neuman MR. 1993. Biomedical sensors. In RC Dorf (Ed.), *The*

Electrical Engineering Handbook,
Boca Raton, FL, CRC Press.

or full-bridge configuration to minimize the temperature sensitivity of the device. Load cells come in various sizes and configurations, and they can measure a wide range of forces.

Measurement of Fluid Dynamic Variables

The measurement of the fluid pressure and flow in both liquids and gases is important in many biomedical applications. These two variables, however, often are the most difficult variables to measure in biologic applications because of interactions with the biologic system and stability problems. Some of the most frequently applied sensors for these measurements are described in the following paragraphs.

Pressure Measurement

Sensors of pressure for biomedical measurements such as blood pressure [7] consist of a structure such as shown in Fig. 9.4. In this case a fluid coupled to the fluid to be measured is housed in a chamber with a flexible diaphragm making up a portion of the wall, with the other side of the diaphragm at atmospheric pressure. When a pressure exists across the diaphragm, it will cause the diaphragm to deflect. This deflection is then measured by a displacement sensor. In the example in Fig. 9.4, the displacement transducer consists of four fine-gauge wires drawn between a structure attached to the diaphragm and the housing of the sensor so that these wires serve as strain gauges. When pressure causes the diaphragm to deflect, two of the fine-wire strain gauges will be extended by a small amount, and the other two will contract by the same amount. By connecting these wires into a bridge circuit, a voltage proportional to the deflection of the diaphragm and hence the pressure can be obtained.

Semiconductor technology has been applied to the design of pressure transducers such that the entire structure can be fabricated from silicon. A portion of a silicon chip can be formed into a diaphragm and semiconductor strain gauges incorporated directly into that diaphragm to produce a small, inexpensive, and sensitive pressure sensor. Such sensors can be used as disposable, single-use devices for measuring blood pressure without the need for additional sterilization before being used on the next patient. This minimizes the risk of transmitting blood-borne infections in the cases where the transducer is coupled directly to the patient's blood for direct blood pressure measurement.

In using this type of sensor to measure blood pressure, it is necessary to couple the chamber containing the diaphragm to the blood or other fluids being measured. This is usually done using a small, flexible plastic tube known as a catheter, that can have one end placed in an artery of the subject while the other is connected to the pressure sensor. This catheter is filled with a physiologic saline solution so that the arterial blood pressure is coupled to the diaphragm. This external blood-pressure-measurement method is used quite frequently in the clinic and research laboratory, but it has the limitation that the properties of the fluid in the catheter and the catheter itself can affect the measurement. For example, both ends of the catheter must be at the same vertical level to avoid a

pressure offset due to hydrostatic effects. Also, the compliance of the tube will affect the frequency response of the pressure measurement. Air bubbles in the catheter or obstructions due to clotted blood or other materials can introduce distortion of the waveform due to resonance and damping. These problems can be minimized by utilizing a miniature semiconductor pressure transducer that is located at the tip of a catheter and can be placed in the blood vessel rather than being positioned external to the body. Such internal pressure sensors are available commercially and have the advantages of a much broader frequency response, no hydrostatic pressure error, and generally clearer signals than the external system.

Although it is possible to measure blood pressure using the techniques described above, this remains one of the major problems in biomedical sensor technology. Long-term stability of pressure transducers is not very good. This is especially true for pressure measurements of venous blood, cerebrospinal fluid, or fluids in the gastrointestinal tract, where pressures are relatively low. Long-term changes in baseline pressure for most pressure sensors require that they be frequently adjusted to be certain of zero pressure. Although this can be done relatively easily when the pressure transducer is located external to the body, this can be a major problem for indwelling pressure transducers. Thus, these transducers must be extremely stable and have low baseline drift to be useful in long-term applications. The packaging of the pressure transducer is also a problem that needs to be addressed, especially when the transducer is in contact with blood for long periods. Not only must the package be biocompatible, but it also must allow the appropriate pressure to be transmitted from the biologic fluid to the diaphragm. Thus, a material that is mechanically stable under corrosive and aqueous environments in the body is needed.

Measurement of Flow

The measurement of true volummetric flow in the body represents one of the most difficult problems in biomedical sensing [8]. The sensors that have been developed measure velocity rather than volume flow, and they can only be used to measure flow if the velocity is measured for a tube of known cross-section. Thus, most flow sensors constrain the vessel to have a specific cross-sectional area.

The most frequently used flow sensor in biomedical systems is the electromagnetic flowmeter illustrated in Fig. 9.5. This device consists of a means of generating a magnetic field transverse to the flow vector in a vessel. A pair of very small biopotential electrodes are attached to the wall of the vessel such that the vessel diameter between them is at right angles to the direction of the magnetic field. As the

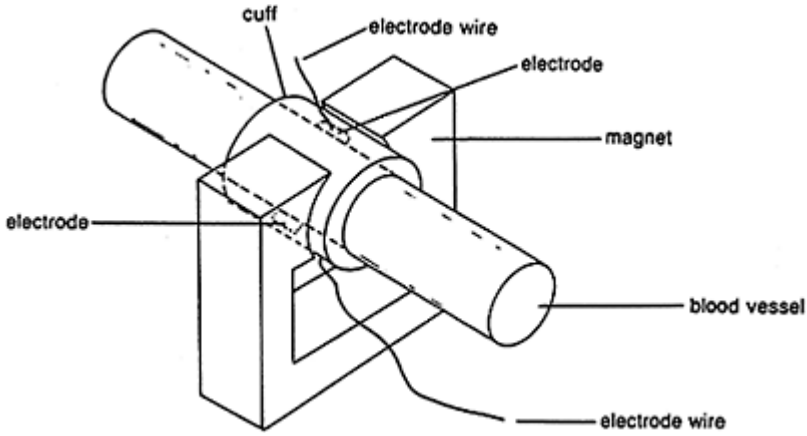


FIGURE 9.5 Fundamental structure of an electromagnetic flowmeter. Reproduced with permission from Neuman MR. 1986. Biosensors: Transducers, electrodes, and physiologic systems. In JD Bronzino (ed), *Biomedical Engineering and Instrumentation: Basic Concepts and Applications*, Boston, PWS Publishers.

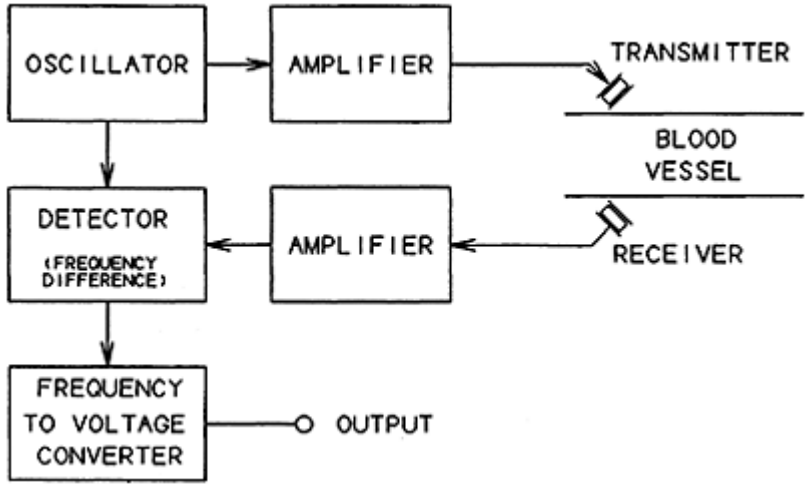


FIGURE 9.6 Structure of an ultrasonic Doppler flowmeter with the major

blocks of the electronic signal processing system. The oscillator generates a signal that, after amplification, drives the transmitting transducer. The oscillator frequency is usually in the range of 1 to 10 MHz. The reflected ultrasound from the blood is sensed by the receiving transducer and amplified before being processed by a detector circuit. This block generates the frequency difference between the transmitted and received ultrasonic signals. This difference frequency can be converted into a voltage proportional to frequency, and hence flow velocity, by the frequency to voltage converter circuit.

blood flows in the structure, ions in the blood deflect in the direction of one or the other electrodes due to the magnetic field, and the voltage across the electrodes is given by

$$v = Blu\,, \tag{9.8}$$

where B is the magnetic field, l is the distance between the electrodes, and u is the average instantaneous velocity of the fluid across the vessel. If the sensor constrains the blood vessel to have a specific diameter, then its cross-sectional area will be known, and multiplying this area by the velocity will give the volume flow.

Although dc flow sensors have been developed and are available commercially, the most desirable method is to use ac excitation of the magnetic field so that offset potential effects from the biopotential electrodes do not generate errors in this measurement.

Small ultrasonic transducers can also be attached to a blood vessel to measure flow as illustrated in Fig. 9.6. In this case the transducers are oriented such that one transmits a continuous ultrasound signal that illuminates the blood. Cells within the blood diffusely reflect this signal in the direction of the second sensor so that the received signal undergoes a Doppler shift in frequency that is proportional to the velocity of the blood. By measuring the frequency shift and knowing the cross-sectional area of the vessel, it is possible to determine the flow.

Another method of measuring flow that has had bio medical application is the measurement of cooling of a heated object by convection. The object is usually a thermistor (see section on temperature measurement, below) placed either in a blood vessel or in tissue, and the thermistor serves as both the heating element and the temperature sensor. In one mode of operation, the amount of power required to maintain

the thermistor at a temperature slightly above that of the blood upstream is measured. As the flow around the thermistor increases, more heat is removed from the thermistor by convection, and so more power is required to keep it at a constant temperature. Relative flow is then measured by determining the amount of power supplied to the thermistor.

In a second approach the thermistor is heated by applying a current pulse and then measuring the cooling curve of the thermistor as the blood flows across it. The thermistor will cool more quickly as the blood flow increases. Both these methods are relatively simple to achieve electronically, but both also have severe limitations. They are essentially qualitative measures and strongly depend on how the ther-

TABLE 9.3 Properties of Temperature Sensors

Sensor	Form	Sensitivity	Stability	Range
Metal resistance thermometer	Coil of fine platinum wire	Low	High	−100–700°C
Thermistor	Bead, disk, or rod	High	Moderate	−50–100°C
Thermocouple	Pair of wires	Low	High	–100–>1000°C
Mercury in glass thermometer	Column of Hg in glass capillary	Moderate	High	−50–400°C
Silicon p-n diode	Electronic component	Moderate	High	−50–150°C

mistor probe is positioned in the vessel being measured. If the probe is closer to the periphery or even in contact with the vessel wall, the measured flow will be different than if the sensor is in the center of the vessel.

Temperature

There are many different sensors of temperature [9], but three find particularly wide application to biomedical problems. Table 9.3 summarizes the properties of various temperature sensors, and these three, including metallic resistance thermometers, thermistors, and thermocouples, are described in the

Metallic Resistance Thermometers

The electric resistance of a piece of metal or wire generally increases as the temperature of that electric conductor increases. A linear approximation to this relationship is given by

$$R = R_0\left[1+\alpha\left(T-T_0\right)\right], \tag{9.9}$$

where R_o is the resistance at temperature T_o, α is the temperature coefficient of resistance, and T is the temperature at which the resistance is being measured. Most metals have temperature coefficients of resistance of the order of 0.1 to 0.4%/°C, as indicated in Table

9.4. The noble metals are preferred for resistance thermometers, since they do not corrode easily and, when drawn into fine wires, their cross-section will remain constant, thus avoiding drift in the resistance over time that could result in an unstable sensor. It is also seen from Table 9.4 that the noble metals of gold and platinum have some of the highest temperature coefficients of resistance of the common metals.

Metal resistance thermometers are often fabricated from fine-gauge insulated wire that is wound into a small coil. It is important in doing so to make certain that there are not other sources of resistance change that could affect the sensor. For example, the structure should be utilized in such a way that no external strains are applied to the wire, since the wire could also behave as a strain gauge. Metallic films and foils can also be used as temperature sensors, and commercial products are available in the wire, foil, or film forms. The electric circuits used to measure resistance, and hence the temperature, are similar to those used with the wire or foil strain gauges. A bridge circuit is the most desirable, although ohmmeter

TABLE 9.4 Temperature Coefficient of Resistance for Common Metals and Alloys

Metal or Alloy	Resistivity at 20°C microhm-cm	Temperature Coefficient of Resistance, %/°C
Platinum	9.83	0.3
Gold	2.22	0.368
Silver	1.629	0.38
Copper	1.724	0.393
Constantan (60% Cu, 40% Ni)	49.0	0.0002
Nichrome (80% Ni, 20% Cr)	108.0	0.013

Source: Pender H, McIlwain K. 1957. Electrical Engineers' Handbook, 4th ed, New York, Wiley.

circuits can also be used. It is important to make sure that the electronic circuit does not pass a large current through the resistance thermometer to provide self-heating due to the Joule conversion of electric energy to heat.

Thermistors

Unlike metals, semiconductor materials have an inverse relationship between resistance and temperature. This characteristic is very nonlinear and cannot be characterized by a linear equation such as for the metals. The thermistor is a semiconductor temperature sensor. Its resistance as a function of temperature is given by

$$R = R_0 e^{\beta\left[\frac{1}{T} - \frac{1}{T_0}\right]}, \tag{9.10}$$

where β is a constant determined by the materials that make up the thermistor. Thermistors can take a variety of forms and cover a large range of resistances. The most common forms used in biomedical applications are the bead, disk, or rod forms of the sensor as illustrated in Fig. 9.7. These structures can be formed from a variety of different semiconductors ranging from elements such as silicon and germanium to mixtures of various semiconducting metallic oxides. Most commercially available thermistors are manufactured from the latter materials, and the specific materials as well as the process for fabricating them are closely held industrial secrets. These materials are chosen not only to have high sensitivity but also to have the greatest stability, since thermistors are generally not as stable as the metallic resistance thermometers. However, thermistors are close to an order of magnitude more sensitive.

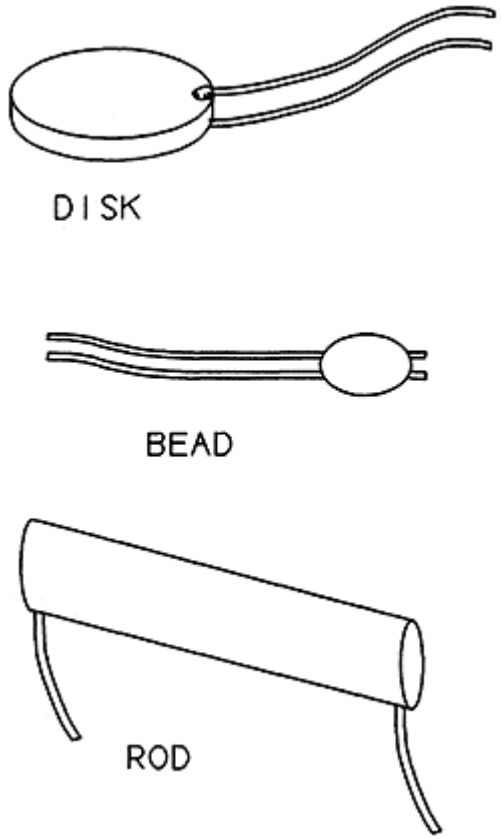


FIGURE 9.7 Common forms of thermistors.

Thermocouples

When different regions of an electric conductor or semiconductor are at different temperatures, there is an electric potential between these regions that is directly related to the temperature differences. This phenomenon, known as the Seebeck effect, can be used to produce a temperature sensor known as *a thermocouple* by taking a wire of metal or alloy *A* and another wire of metal or alloy *B* and connecting them as shown in Fig. 9.8. One of the junctions is known as the sensing junction, and the other is the reference junction. When these junctions are at different temperatures, a voltage proportional to the temperature difference will be seen at the voltmeter when metals *A* and *B* have different Seebeck coefficients. This voltage is roughly proportional to the temperature difference and can be represented over the relatively small temperature differences encountered in biomedical applications by the linear equation,

$$V = S_{AB}\,(T_s - T_r), \tag{9.11}$$

where S_{AB} is the Seebeck coefficient for the thermocouple made up of metals *A* and *B*. Although this equation is a reasonable approximation, more accurate data are usually found in tables of actual voltages as a function of temperature difference. In some applications the voltmeter is located at the reference junction, and one uses some independent means such as a mercury in glass thermometer to measure the reference junction temperature. Where precision measurements are made, the reference junction is often placed in an environment of known temperature such as an ice bath. Electronic measurement of reference junction temperature can also be carried out and used to compensate for the reference junction temperature so that the voltmeter reads a signal equivalent to what would be seen if the reference junction were at 0°C. This electronic reference junction compensation is usually carried out using a metal resistance temperature sensor to determine reference junction temperature.

The voltages generated by thermocouples used for temperature measurement are generally quite small being on the order of tens of microvolts per degree C. Thus, for most biomedical measurements where there is only a small difference in temperature between the sensing and reference junction, very sensitive amplifiers must be used to measure these potentials. Thermocouples have been used in industry for temperature measurement for many years. Several standard alloys to provide optimal sensitivity and stability of these sensors have evolved. Table 9.5 lists these common alloys, the Seebeck coefficient for thermocouples of these materials at room temperature, and the full range of temperatures over which these thermocouples can be used.

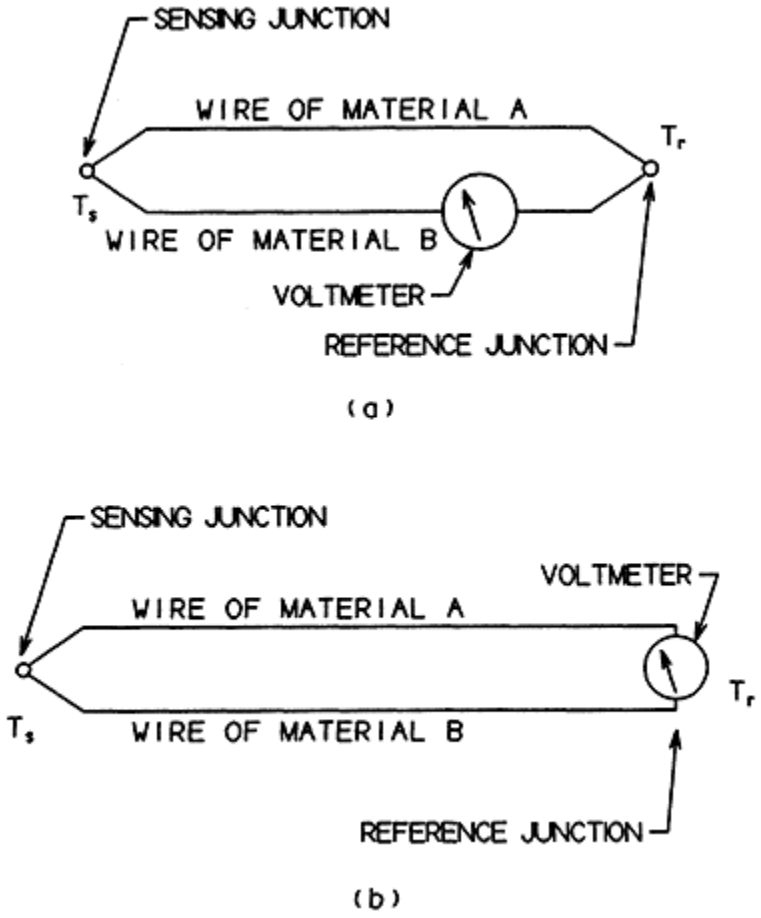


FIGURE 9.8 Circuit arrangement for a thermocouple showing the voltage-measuring device, the voltmeter, interrupting one of the thermocouple wires *(a)* and at the cold junction *(b)*.

TABLE 9.5 Common Thermocouples

Type	Materials	Seebeck Coefficient, μV/°C	Temperature Range
S	Platinum/platinum 10% rhodium	6	0–1700°C
T	Copper/constantan	50	−190–400°C
K	Chromel/alumel	41	−200–1370°C
J	Iron/constantan	53	−200–760°C

E	Chromel/constantan	78	−200–970°C

Thermocouples can be fabricated in many different ways depending on their applications. They are especially suitable for measuring temperature differences between two structures, since the sensing junction can be placed on one while the other has the reference junction. Higher-output thermocouples or thermopiles can be produced by connecting several thermocouples in series. Thermocouples can be made from very fine wires that can be implanted in biologic tissues for temperature measurements, and it is also possible to place these fine-wire thermocouples within the lumen of a hypodermic needle to make short-term temperature measurements in tissue.

9.2 Biomedical Applications of Physical Sensors

Just as it is not possible to cover the full range of physical sensors in this chapter, it is also impossible to consider the many biomedical applications that have been reported for these sensors. Instead, some representative examples will be given. These are summarized in Table 9.6 and will be briefly described in the following paragraphs.

Liquid metal strain gauges are especially useful in biomedical applications, because they are mechanically compliant and provide a better mechanical impedance match to most biomedical tissues than other types of strain gauges. By wrapping one of these strain gauges around a circumference of the abdomen, it will stretch and contract with the abdominal breathing movements. The signal from the strain gauge can then be used to monitor breathing in patients or experimental animals. The advantage of this sensor is its compliance so that it does not interfere with the breathing movements or substantially increase the required breathing effort.

One of the original applications of the liquid metal strain gauge was in limb plethysmography [3]. One or more of these sensors are wrapped around an arm or leg at various points and can be used to measure changes in circumference that are related to the cross-sectional area and hence the volume of the limb at those points. If the venous drainage from the limb is occluded, the limb volume will increase as it fills with blood. Releasing the occlusion allows the volume to return to normal. The rate of this decrease in volume can be monitored using the liquid metal strain gauges, and this can be used to identify venous blockage when the return to baseline volume is too slow.

Breathing movements, although not volume, can be seen using a simple magnetic velocity detector. By placing a small permanent magnet on the anterior side of the chest or abdomen and a flat, large-area coil on the posterior side opposite from the magnet, voltages are induced in the coil as the chest of abdomen moves during breathing. The voltage itself can be used to detect the presence of breathing movements, or it can be electronically integrated to give a signal related to displacement.

The LVDT is a displacement sensor that can be used for more precise applications. For example, it can be used in studies of muscle physiology to measure the displacement of a muscle or to measure the isometric force generated by the muscle (using a load cell) and must ensure that there is no muscle

TABLE 9.6 Examples of Biomedical Applications of Physical Sensors

Sensor	Application	Signal Range	Reference
Liquid metal strain gauge	Breathing movement	0–0.05	
	Limb plethysmography	0–0.02	3
Magnetic displacement sensor	Breathing movement	0–10 mm	10
LVDT	Muscle contraction	0–20 mm	
	Uterine contraction sensor	0–5 mm	11
Load cell	Electronic scale	0–440 lbs (0–200 kg)	12
Accelerometer	Subject activity	0–20 m/s^2	13
Miniature silicon pressure sensor	Intra- arterial blood pressure	0–50 Pa (0–350 mm Hg)	
	Urinary bladder pressure	0–10 Pa (0–70 mm Hg)	
	Intrauterine pressure	0–15 Pa (0–100 mm Hg)	14
Electromagnetic flow sensor	Cardiac output (with integrator)	0–500 ml/min	
	Organ blood flow	0–100 ml/min	15

movement. It can also be incorporated into other physical sensor such as a pressure sensors or a tocodynamometer, a sensor used to electronically “feel” uterine contractions of patients in labor or those at risk of premature labor and delivery.

In addition to studying muscle forces, load cells can be used in various types of electronic scales for weighing patients or study animals. The simplest electronic scale consists of a platform placed on top of a load cell. The weight of any object placed on the platform will produce a force that can be sensed by the load cell. In some critical care situations in the hospital, it is important to carefully monitor the weight of a patient. For example, this is important in watching water balance in patients receiving fluid therapy. The electronic scale concept can be extended by placing a load cell under each leg of the patient’s bed and summing the forces seen by each load cell to get the total weight of the patient and the bed. Since the bed weight remains fixed, weight changes seen will reflect changes in patient weight.

Accelerometers can be used to measure patient or research subject activity. By attaching a small accelerometer to the individual being studied, any movements can be detected. This can be useful in sleep studies where movement can help to determine the sleep state. Miniature accelerometers and recording devices can also be worn by patients to study activity patterns and determine effects of disease or treatments on patient activity [13].

Miniature silicon pressure sensors are used for the indwelling measurement of fluid pressure in most body cavities. The measurement of intra-arterial blood pressure is the most frequent application, but pressures in other cavities such as the urinary bladder and the uterus are also measured. The small size of these sensors and the resulting ease of introduction of the sensor into the cavity make these sensors important for these applications.

The electromagnetic flow sensor has been a standard method in use in the physiology laboratory for many years. Its primary application has been for measurement of cardiac output and blood flow to specific organs in research animals. New miniature inverted electromagnetic flow sensors make it possible to temporarily introduce a flow probe into an artery through its lumen to make clinical measurements.

The measurement of body temperature using instruments employing thermistors as the sensor has greatly increased in recent years. Rapid response times of these low-mass sensors make it possible to quickly assess patients' body temperatures so that more patients can be evaluated in a given period. This can then help to reduce healthcare costs. The rapid response time of low-mass thermistors makes them a simple sensor to be used for sensing breathing. By placing small thermistors near the nose and mouth, the elevated temperature of exhaled air can be sensed to document a breath.

The potential applications of physical sensors in medicine and biology are almost limitless. To be able to use these devices, however, scientists must first be familiar with the underlying sensing principles. It is then possible to apply these in a form that addresses the problems at hand.

References

1. Doebelin EO. 1990. *Measurement Systems: Applications and Design*, New York, McGraw-Hill.
2. Dechow PC. 1988. Strain Gauges. In J Webster (ed), *Encyclopedia of Medical Devices and Instrumentation*, pp 2715–2721, New York, Wiley.
3. Whitney RJ. 1949. The measurement of changes in human limb-volume by means of a mercury-in-rubber strain gauge. *J Physiol* 109:5p.
4. Schaevitz H. 1947. The linear variable differential transformer. *Proc Soc Stress Anal* 4:79.
5. Stegall HF, Kardon MB, Stone HL, et al. 1967. A portable simple sonomicrometer. *J Appl Physiol* 23:289.
6. Angelsen BA, Brubakk AO. 1976. Transcutaneous measurement of blood flow velocity in the human aorta. *Cardiovasc Res* 10:368.
7. Geddes LA. 1970. *The Direct and Indirect Measurement of Blood Pressure*, Chicago, Year Book.
8. Roberts VC. 1972. *Blood Flow Measurements*, Baltimore, Williams & Wilkins.
9. Herzfeld CM (ed). 1962. *Temperature: Its Measurement and Control in Science and Industry*, New York, Reinhold.
10. Rolfe P. 1971. A magnetometer respiration monitor for use with premature babies. *Biomed Eng* 6:402.
11. Reddy NP, Kesavan SK. 1988. Linear variable differential transformers. In J Webster (ed), *Encyclopedia of Medical Devices and Instrumentation*, pp 1800–1806, New York, Wiley.
12. Roe FC. 1966. New equipment for metabolic studies. *Nurs Clin N Am* 1:621.
13. Patterson SM, Krantz DS, Montgomery LC, et al. 1993. Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology* 30:296.

14. Fleming DG, Ko WH, Neuman MR (eds). 1977. *Indwelling and Implantable Pressure Transducers*, Cleveland, CRC Press.
15. Wyatt DG. 1971. Electromagnetic blood flow measurements. In BW Watson (ed), *IEE Medical Electronics Monographs*, London, Peregrinus.

Further Information

Good overviews of physical sensors are found in these books: Doebelin EO. 1990. *Measurement Systems: Application and Design*, 4th ed, New York, McGraw-Hill; Harvey, GF (Ed). 1969. *Transducer Compendium*, 2d ed, New York, Plenum. One can also find good descriptions of physical sensors in chapters of two works edited by John Webster. Chapters 2, 7, and 8 of his textbook (1992) *Medical Instrumentation: Application and Design*, 2d ed, Boston, Houghton Mifflin, and several articles in his *Encyclopedia on Medical Devices and Instrumentation*, published by Wiley in 1988, cover topics on physical sensors.

Although a bit old, the text, *Transducers for Biomedical Measurements* (New York, J.Wiley, 1974), by Richard S.C.Cobbold, remains one of the best descriptions of biomedical sensors available. By supplementing the material in this book with recent manufacturers' literature, the reader can obtain a wealth of information on physical (and for that matter chemical) sensors for biomedical application.

The journals *IEEE Transactions on Biomedical Engineering* and *Medical and Biological Engineering and Computing* are good sources of recent research on biomedical applications of physical sensors. The journals *Physiological Measurement* and *Sensors and Actuators* are also good sources for this material.

10
Biopotential Electrodes

Michael R.Neuman
Joint Program in Biomedical Engineering, The University of Memphis and University of Tennessee, Memphis

Biologic systems frequently have electric activity associated with them. This activity can be a constant dc electric field, a constant flux of charge-carrying particles or current, or a time-varying electric field or current associated with some time-dependent biologic or biochemical phenomenon. Bioelectric phenomena are associated with the distribution of ions or charged molecules in a biologic structure and the changes in this distribution resulting from specific processes. These changes can occur as a result of biochemical reactions, or they can emanate from phenomena that alter local anatomy.

One can find bioelectric phenomena associated with just about every organ system in the body. Nevertheless, a large proportion of these signals are associated with phenomena that are at the present time not especially useful in clinical medicine and represent time-invariant, low-level signals that are not easily measured in practice. There are, however, several signals that are of diagnostic significance or that provide a means of electronic assessment to aid in understanding biologic systems. These signals, their usual abbreviations, and the systems they measure are listed in Table 10.1. Of these, the most familiar is the electrocardiogram, a signal derived from the electric activity of the heart. This signal is widely used in diagnosing disturbances in cardiac rhythm, signal conduction through the heart, and damage due to cardiac ischemia and infarction. The electromyogram is used for diagnosing neuromuscular diseases, and the electroencephalogram is important in identifying brain dysfunction and evaluating sleep. The other signals listed in Table 10.1 are currently of lesser diagnostic significance but are, nevertheless, used for studies of the associated organ systems.

Although Table 10.1 and the above discussion are concerned with bioelectric phenomena in animals and these techniques are used primarily in studying mammals, bioelectric signals also arise from plants [1]. These signals are generally steady-state or slowly changing, as opposed to the time-varying signals listed in Table 10.1. An extensive literature exists on the origins of bioelectric signals, and the interested reviewer is referred to the text by Plonsey and Barr for a general overview of this area [2],

0-8493-1813-0/03/$0.00+$1.50

10.1 Sensing Bioelectric Signals

The mechanism of electric conductivity in the body involves ions as charge carriers. Thus, picking up bioelectric signals involves interacting with these ionic charge carriers and transducing ionic currents into electric currents required by wires and electronic instrumentation. This transducing function is carried out by electrodes that consist of electrical conductors in contact with the aqueous ionic solutions

TABLE 10.1 Bioelectric Signals Sensed by Biopotential Electrodes and Their Sources

Bioelectric Signal	Abbreviation	Biologic Source
Electrocardiogram	EGG	Heart—as seen from body surface
Cardiac electrogram	—	Heart—as seen from within
Electromyogram	EMG	Muscle
Electroencephalogram	EEC	Brain
Electrooptigram	EOG	Eye dipole field
Electroretinogram	ERG	Eye retina
Action potential	—	Nerve or muscle
Electrogastrogram	EGG	Stomach
Galvanic skin reflex	GSR	Skin

of the body. The interaction between electrons in the electrodes and ions in the body can greatly affect the performance of these sensors and requires that specific considerations be made in their application.

At the interface between an electrode and an ionic solution redox (oxidation-reduction), reactions need to occur for a charge to be transferred between the electrode and the solution. These reactions can be represented in general by the following equations:

$$C \rightleftharpoons C^{n+} + ne^{-} \tag{10.1}$$

$$A^{m-} \rightleftharpoons A + me^{-}, \tag{10.2}$$

where n is the valence of cation material C, and m is the valence of anion material, A. For most electrode systems, the cations in solution and the metal of the electrodes are the same, so the atoms C are oxidized when they give up electrons and go into solution as positively charged ions. These ions are reduced when the process occurs in the reverse direction. In the case of the anion reaction, Eq. (10.2), the directions for oxidation and

reduction are reversed. For best operation of the electrodes, these two reactions should be reversible, that is, it should be just as easy for them to occur in one direction as the other.

The interaction between a metal in contact with a solution of its ions produces a local change in the concentration of the ions in solution near the metal surface. This causes charge neutrality not to be maintained in this region, causing the electrolyte surrounding the metal to be at a different electrical potential from the rest of the solution. Thus, a potential difference known as the *half-cell potential* is established between the metal and the bulk of the electrolyte. It is found that different characteristic potentials occur for different materials, and some of these potentials are summarized in Table 10.2. These half-cell potentials can be important when using electrodes for low frequency or dc measurements.

The relationship between electric potential and ionic concentrations or, more precisely, ionic activities is frequently considered in electrochemistry. Most commonly two ionic solutions of different activity are separated by an ion-selective semipermeable membrane that allows one type of ion to pass freely through

TABLE 10.2 Half-cell Potentials for Materials and Reactions Encountered in Biopotential Measurement

Metal and Reaction	Half-Cell Potential, V
$A1 \rightarrow Al^{3}+3e^{-}$	–1.706
$Ni \rightarrow Ni^{2+}+2e^{-}$	−0.230
$H_2 \rightarrow 2H^{+}+2e^{-}$	0.000 (by definition)
$Ag+Cl^{-} \rightarrow +e^{-}$	+0.223
$Ag \rightarrow Ag^{+}+e^{-}$	+0.799
$Au \rightarrow Au^{+}+e^{-}$	+1.680

the membrane. It can be shown that an electric potential E will exist between the solutions on either side of the membrane, based upon the relative activity of the permeable ions in each of these solutions. This relationship is known as the Nernst equation,

$$E = -\frac{RT}{nF} \ln\left(\frac{a_1}{a_2}\right), \tag{10.3}$$

where a_1 and a_2 are the activities of the ions on either side of the membrane, R is the universal gas constant, T is the absolute temperature, n is the valence of the ions, and F is the Faraday constant. More detail on this relationship can be found in Chapter 11.

When no electric current flows between an electrode and the solution of its ions or across an ion-permeable membrane, the potential observed should be the half-cell potential or the Nernst potential, respectively. If, however, there is a current, these potentials can be altered. The difference between the potential at zero current and the

measured potentials while current is passing is known as the *overvoltage* and is the result of an alteration in the charge distribution in the solution in contact with the electrodes or the ion-selective membrane. This effect is known as polarization and can result in diminished electrode performance, especially under conditions of motion. There are three basic components to the polarization overpotential: the ohmic, the concentration, and the activation overpotentials. Of these, the activation overpotential is of greatest concern in bioelectric measurements. More details on these overpotentials can be found in electrochemistry or biomedical instrumentation texts [4].

Perfectly polarizable electrodes pass a current between the electrode and the electrolytic solution by changing the charge distribution within the solution near the electrode. Thus, no actual current crosses the electrode-electrolyte interface. Nonpolarized electrodes, however, allow the current to pass freely across the electrode-electrolyte interface without changing the charge distribution in the electrolytic solution adjacent to the electrode. Although these types of electrodes can be described theoretically, neither can be fabricated in practice. It is possible, however, to come up with electrode structures that closely approximate their characteristics.

Electrodes made from noble metals such as platinum are often highly polarizable. A charge distribution different from that of the bulk electrolytic solution is found in the solution close to the electrode surface. Such a distribution can create serious limitations when movement is present and the measurement involves low frequency or even dc signals. If the electrode moves with respect to the electrolytic solution, the charge distribution in the solution adjacent to the electrode surface will change, and this will induce a voltage change in the electrode that will appear as motion artifact in the measurement. Thus, for most biomedical measurements, nonpolarizable electrodes are preferred to those that are polarizable.

The silver-silver chloride electrode is one that has characteristics similar to a perfectly nonpolarizable electrode and is practical for use in many biomedical applications. The electrode (Fig. 10.1*a*) consists of a silver base structure that is coated with a layer of the ionic compound silver chloride. Some of the silver chloride when exposed to light is reduced to metallic silver, so a typical silver—silver chloride electrode has finely divided metallic silver within a matrix of silver chloride on its surface. Since the silver chloride is relatively insoluble in aqueous solutions, this surface remains stable. Because there is minimal polarization associated with this electrode, motion artifact is reduced compared to polarizable electrodes such as the platinum electrode. Furthermore, due to the reduction in polarization, there is also a smaller effect of frequency on electrode impedance, especially at low frequencies.

Silver-silver chloride electrodes of this type can be fabricated by starting with a silver base and electrolytically growing the silver chloride layer on its surface [4]. Although an electrode produced in this way can be used for most biomedical measurements, it is not a robust structure, and pieces of the silver chloride film can be chipped away after repeated use of the structure. A structure with greater mechanical stability is the sintered silver-silver chloride electrode in Fig. 10.1*b*. This electrode consists of a silver lead wire surrounded by a sintered cylinder made up of finely divided silver and silver-chloride powder pressed together.

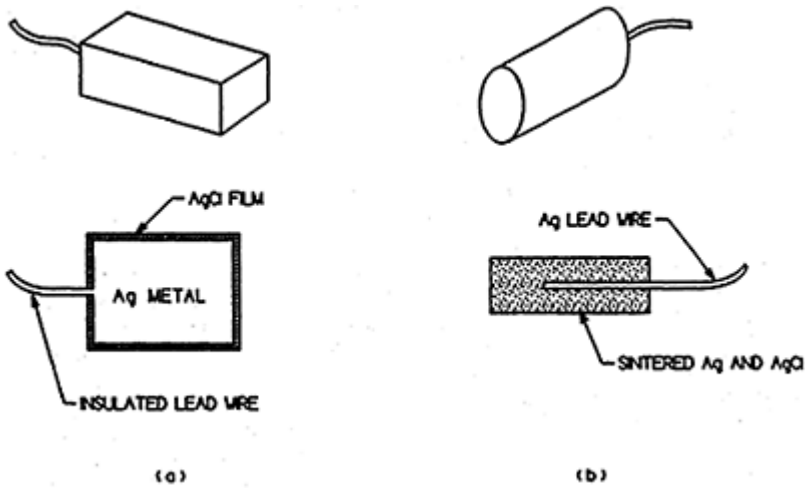


FIGURE 10.1 Silver-silver electrodes for biopotential measurements: *(a)* metallic silver with a silver chloride surface layer and *(b)* sintered electrode structure. The lower views show the electrodes in cross-section.

In addition to its nonpolarizable behavior, the silver-silver chloride electrode exhibits less electrical noise than the equivalent polarizable electrodes. This is especially true at low frequencies, and so silver-silver chloride electrodes are recommended for measurements involving very low voltages for signals that are made up primarily of low frequencies. A more detailed description of silver-silver chloride electrodes and methods to fabricate these devices can be found in Janz and Ives [5] and biomedical instrumentation textbooks [4].

10.2 Electrical Characteristics

The electrical characteristics of biopotential electrodes are generally nonlinear and a function of the current density at their surface. Thus, having the devices represented by linear models requires that they be operated at low potentials and currents. Under these idealized conditions, electrodes can be represented by an equivalent circuit of the form shown in Fig. 10.2. In this circuit R_d and C_d are components that represent the impedance associated with the electrode-electrolyte interface and polarization at this interface. Rs is the series resistance associated with interfacial effects and the resistance of the electrode materials

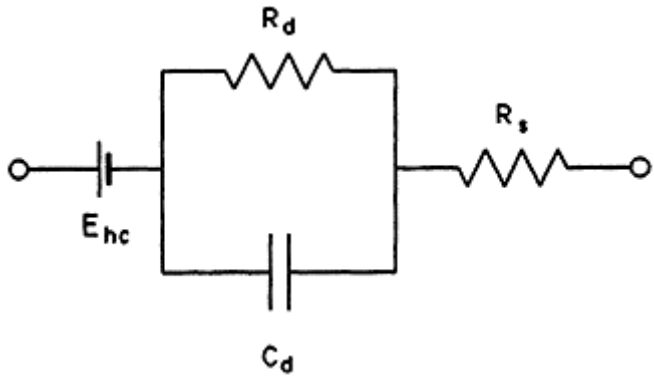


FIGURE 10.2 The equivalent circuit for a biopotential electrode.

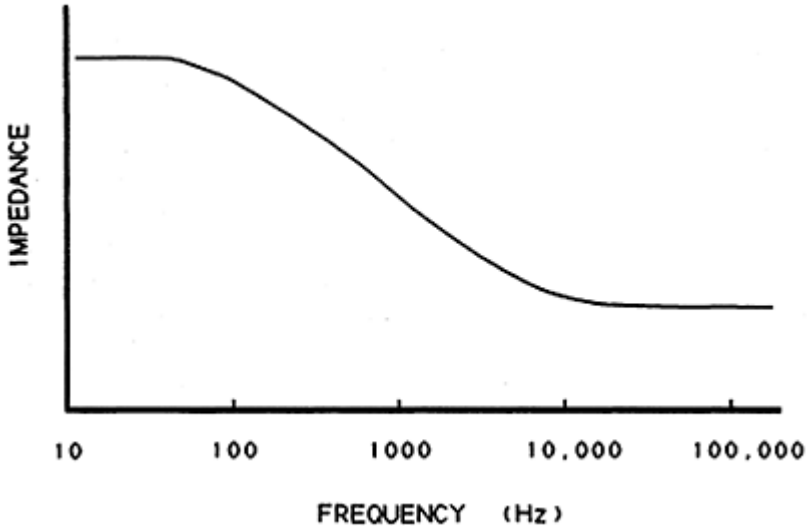


FIGURE 10.3 An example of biopotential electrode impedance as a function of frequency. Characteristic frequencies will be somewhat different for electrode different geometries and materials.

TABLE 10.3 The Effect of Electrode Properties on Electrode Impedance

Property	Change in Property	Changes in Electrode Impedance
Surface area	≠	Ø
Polarization	≠	≠At low frequencies
Surface roughness	≠	Ø
Radius of curvature	≠	Ø
Surface contamination	≠	≠

themselves. The battery E_{hc} represents the half-cell potential described above. It is seen that the impedance of this electrode will be frequency dependent, as illustrated in Fig. 10.3. At low frequencies the impedance is dominated by the series combination of R_s and R_d, whereas at higher frequencies C_d bypasses the effect of R_d so that the impedance is now close to R_s. Thus, by measuring the impedance of an electrode at high and low frequencies, it is possible to determine the component values for the equivalent circuit for that electrode.

The electrical characteristics of electrodes are affected by many physical properties of these electrodes. Table 10.3 lists some of the more common physical properties of electrodes and qualitatively indicates how these can affect electrode impedance.

10.3 Practical Electrodes for Biomedical Measurements

Many different forms of electrodes have been developed for different types of biomedical measurements. To describe each of these would go beyond the constraints of this chapter, but some of the more commonly used electrodes are presented in this section. The reader is referred to the monograph by Geddes for more details and a wider selection of practical electrodes [6].

Body-Surface Biopotential Electrodes

This category includes electrodes that can be placed on the body surface for recording bioelectric signals. The integrity of the skin is not compromised when these electrodes are applied, and they can be used for short-term diagnostic recording such as taking a clinical electrocardiogram or long-term chronic recording such as occurs in cardiac monitoring.

Metal Plate Electrodes

The basic metal plate electrode consists of a metallic conductor in contact with the skin with a thin layer of an electrolyte gel between the metal and the skin to establish this contact. Examples of metal plate electrodes are seen in Fig. 10.4*a*. Metals commonly

used for this type of electrode include German silver (a nickel-silver alloy), silver, gold, and platinum. Sometimes these electrodes are made of a foil of the metal so as to be flexible, and sometimes they are produced in the form of a suction electrode (Fig. 10.4*b*) to make it easier to attach the electrode to the skin to make a measurement and then move it to another point to repeat the measurement. These types of electrodes are used primarily for diagnostic recordings of biopotentials such as the electrocardiogram or the electroencephalogram. Metal disk electrodes with a gold surface in a conical shape such as shown in Fig. 10.4*c* are frequently used for EEC recordings. The apex of the cone is open so that electrolyte gel or paste can be introduced to both make good contact between the electrode and the head and to allow this contact medium to be replaced should it dry out during its use.

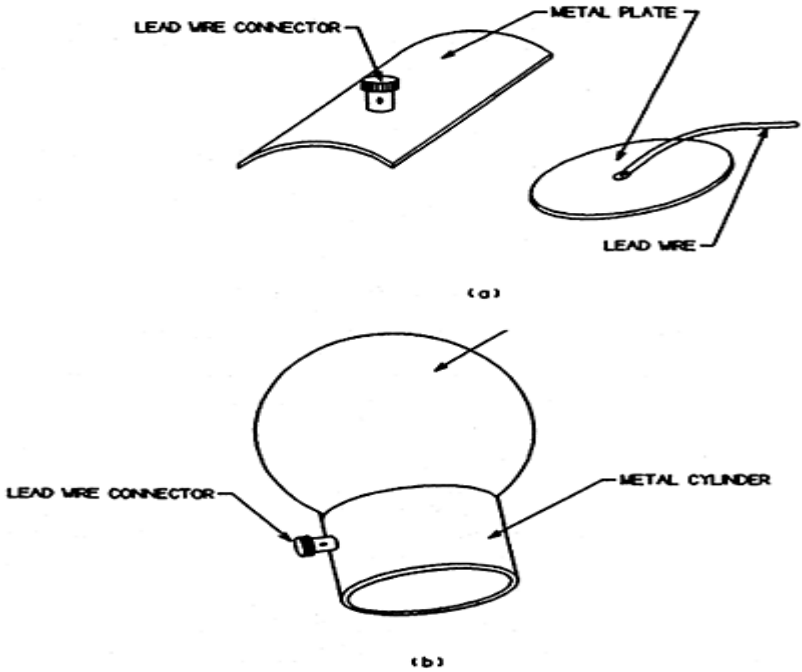
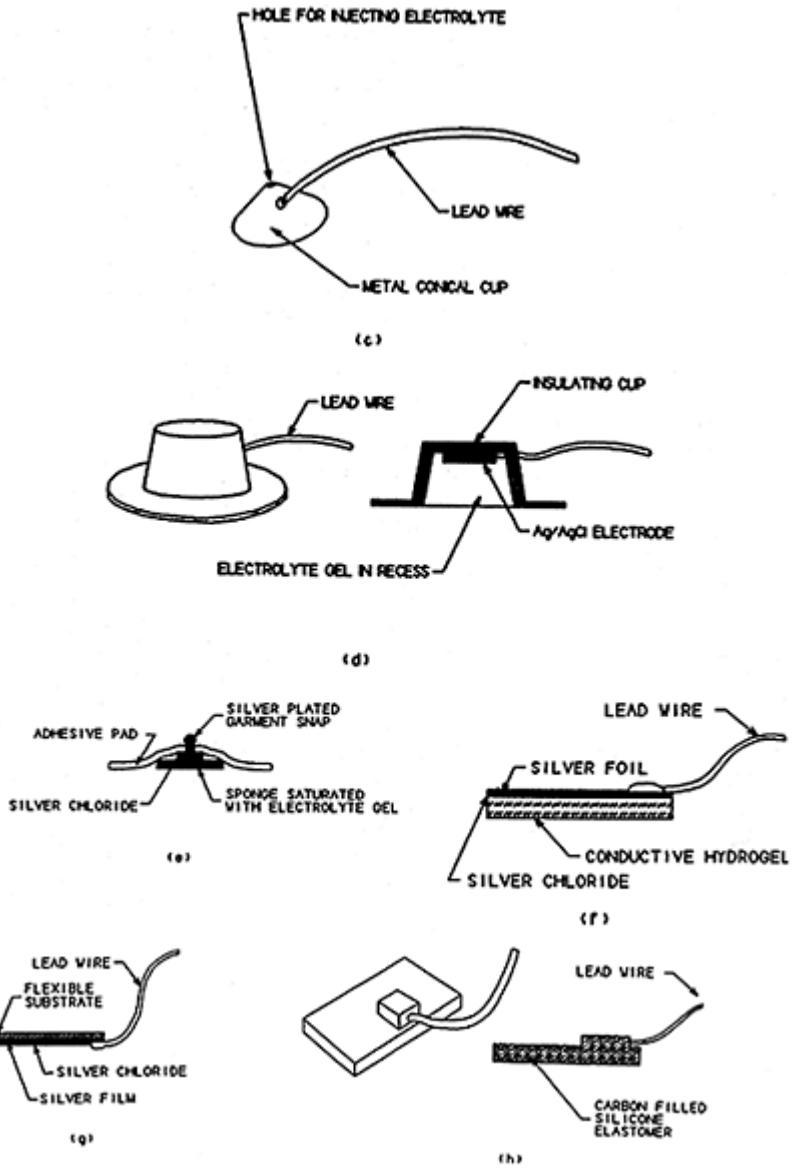


FIGURE 10.4 Examples of different skin electrodes: *(a)* metal plate electrodes, *(b)* suction electrode for EGG, *(c)* metal cup EEG electrode, *(d)* recessed electrode, *(e)* disposable electrode with electrolyte-impregnated sponge (shown in cross-section), (f) disposable hydrogel electrode (shown in cross-section), *(g)* thin-film electrode for use with neonates (shown

in cross-section), *(h)* carbon-filled elastomer dry electrode.



Electrodes for Chronic Patient Monitoring

Long-term monitoring of biopotentials such as the electrocardiogram as performed by cardiac monitors places special constraints on the electrodes used to pick up the signals.

These electrodes must have a stable interface between them and the body, and frequently nonpolarizable electrodes are, therefore, the best for this application. Mechanical stability of the interface between the electrode and the skin can help to reduce motion artifact, and so there are various approaches to reduce interfacial motion between the electrode and the coupling electrolyte or the skin. Figure 10.4*d* is an example of one approach to reduce motion artifact by recessing the electrode in a cup of electrolytic fluid or gel. The cup is then securely fastened to the skin surface using a double-sided adhesive ring. Movement of the skin with respect to the electrode may affect the electrolyte near the skin-electrolyte interface, but the electrode-electrolyte interface can be several millimeters away from this location, since it is recessed in the cup. The fluid movement is unlikely to affect the recessed electrode-electrolyte interface as compared to what would happen if the electrode was separated from the skin by just a thin layer of electrolyte.

The advantages of the recessed electrode can be realized in a simpler design that lends itself to mass production through automation. This results in low per-unit cost so that these electrodes can be considered disposable. Figure 10.4*e* illustrates such an electrode in cross-section. The electrolyte layer now consists of an open-celled sponge saturated with a thickened (high-viscosity) electrolytic solution. The sponge serves the same function as the recess in the cup electrodes and is coupled directly to a silver-silver chloride electrode. Frequently, the electrode itself is attached to a clothing snap through an insulating-adhesive disk that holds the structure against the skin. This snap serves as the point of connection to a lead wire. Many commercial versions of these electrodes in various sizes are available, including electrodes with a silver-silver chloride interface or ones that use metallic silver as the electrode material.

A recently developed modification of this basic monitoring electrode structure is shown in Fig. 10.4*f*. In this case the metal electrode is a silver foil with a surface coating of silver chloride. The foil gives the electrode increased flexibility to fit more closely over body contours. Instead of using the sponge, a hydrogel film (really a sponge on a microscopic level) saturated with an electrolytic solution and formed from materials that are very sticky is placed over the electrode surface. The opposite surface of the hydrogel layer can be attached directly to the skin, and since it is very sticky, no additional adhesive is needed. The mobility and concentration of ions in the hydrogel layer is generally lower than for the electrolytic solution used in the sponge or the cup. This results in an electrode that has a higher source impedance as compared to these other structures. An important advantage of this structure is its ability to have the electrolyte stick directly on the skin. This greatly reduces interfacial motion between the skin surface and the electrolyte, and hence there is a smaller amount of motion artifact in the signal. This type of hydrogel electrode is, therefore, especially valuable in monitoring patients who move a great deal or during exercise.

Thin-film flexible electrodes such as shown in Fig. 10.4*g* have been used for monitoring neonates. They are basically the same as the metal plate electrodes; only the thickness of the metal in this case is less than a micrometer. These metal films need to be supported on a flexible plastic substrate such as polyester or polyimide. The advantage of using only a thin metal layer for the electrode lies in the fact that these electrodes will then become x-ray transparent. This is especially important in infants where repeated placement and removal of electrodes, so that x-rays may be taken, can cause substantial skin irritation.

Electrodes that do not use artificially applied electrolyte solutions or gels and, therefore, are often referred to as dry electrodes have been used in some monitoring applications. These sensors, as illustrated in Fig. 10.4*h*, can be placed on the skin and held in position by an elastic band or tape. They are made up of a graphite or metal-filled polymer such as silicone. The conducting particles are ground into a fine powder, and this is added to the silicone elastomer before it cures to produce a conductive material with physical properties similar to that of the elastomer. When held against the skin surface, these electrodes establish contact with the skin without the need for an electrolytic fluid or gel. In actuality such a layer is formed by sweat under the electrode surface. For this reason these electrodes tend to perform better after they have been left in place for an hour or two so that this layer forms. Some investigators have found that placing a drop of physiologic saline solution on the skin before applying the electrode accelerates this process. This type of electrode has found wide application in home infant cardiorespiratory monitoring because of the ease with which it can be applied by untrained caregivers.

Intracavitary and Intratissue Electrodes

Electrodes can be placed within the body for biopotential measurements. These electrodes are generally smaller than skin surface electrodes and do not require special electrolytic coupling fluid, since natural body fluids serve this function. There are many different designs for these internal electrodes, and only a few examples are given in the following paragraphs. Basically these electrodes can be classified as needle electrodes, which can be used to penetrate the skin and tissue to reach the point where the measurement is to be made, or they are electrodes that can be placed in a natural cavity or surgically produced cavity in tissue. Figure 10.5 illustrates some of these internal electrodes.

A catheter tip or probe electrode is placed in a naturally occurring cavity in the body such as in the gastrointestinal system. A metal tip or segment on a catheter makes up the electrode. The catheter or, in the case where there is no hollow lumen, probe, is inserted into the cavity so that the metal electrode makes contact with the tissue. A lead wire down the lumen of the catheter or down the center of the probe connects the electrode to the external circuitry.

The basic needle electrode shown in Fig. 10.5*b* consists of a solid needle, usually made of stainless steel, with a sharp point. An insulating material coats the shank of the needle up to a millimeter or two of the tip so that the very tip of the needle remains exposed. When this structure is placed in tissue such as skeletal muscle, electrical signals can be picked up by the exposed tip. One can also make needle electrodes by running one or more insulated wires down the lumen of a standard hypodermic needle. The electrode, as shown in Fig. 10.5*c*, is shielded by the metal of the needle and can be used to pick up very localized signals in tissue.

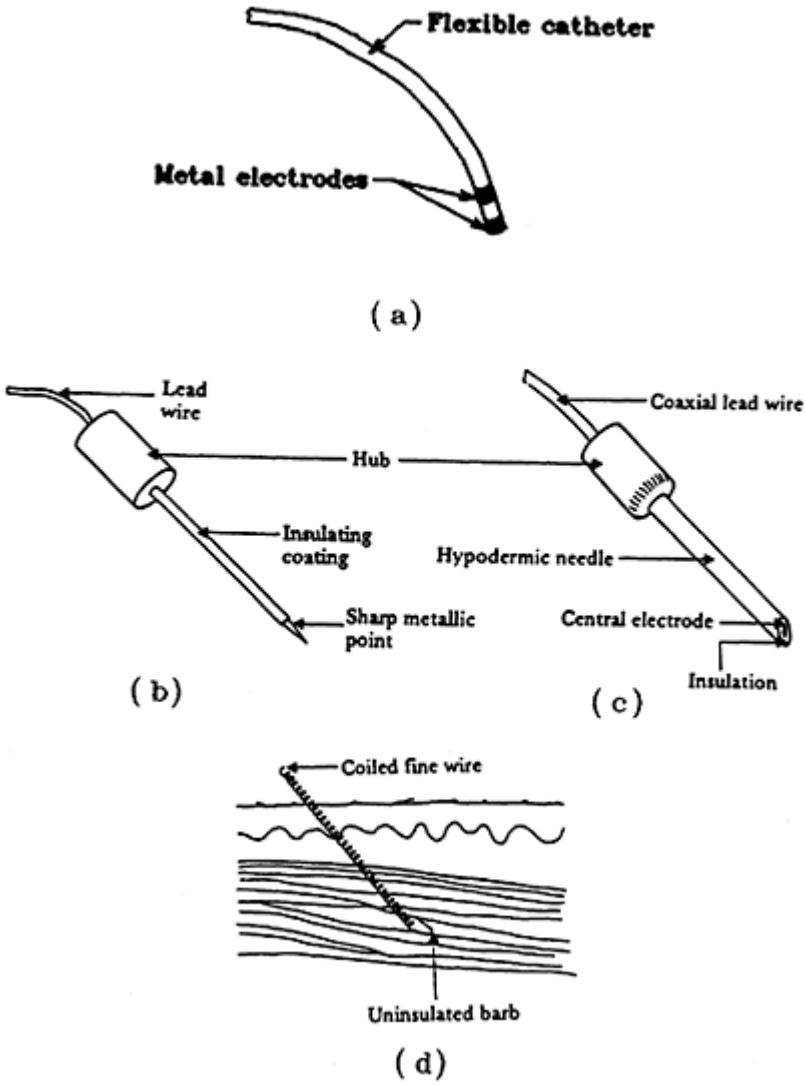


FIGURE 10.5 Examples of different internal electrodes: *(a)* catheter or probe electrode, *(b)* needle electrode, *(c)* coaxial needle electrode, *(d)* coiled wire electrode. (Reprinted with permission from Webster JG (Ed). 1992. *Medical Instrumentation: Application and Design*, Houghton Mifflin, Boston.)

Fine wires can also be introduced into tissue using a hypodermic needle, which is then withdrawn. This wire can remain in tissue for acute or chronic measurements. Caldwell and Reswick have used fine coiled wire electrodes in skeletal muscle for several years without adverse effects [7].

Microelectrodes

The electrodes described in the previous paragraphs have been applied to studying bioelectric signals at the organism, organ, or tissue level but not at the cellular level. To study the electric behavior of cells, electrodes that are themselves smaller than the cells being studied need to be used. Three types of electrodes have been described for this purpose: etched metal electrodes, micropipette electrodes, and metal-film-coated micropipette electrodes. The metal microelectrode is essentially a subminiature version of the needle electrode described in the previous section (Fig. 10.6*a*). In this case, a strong metal such as tungsten is used. One end of this wire is etched electrolytically to give tip diameters on the order of a few micrometers. The structure is insulated up to its tip, and it can be passed through the membrane of a cell to contact the cytosol. The advantage of these electrodes is that they are both small and robust and

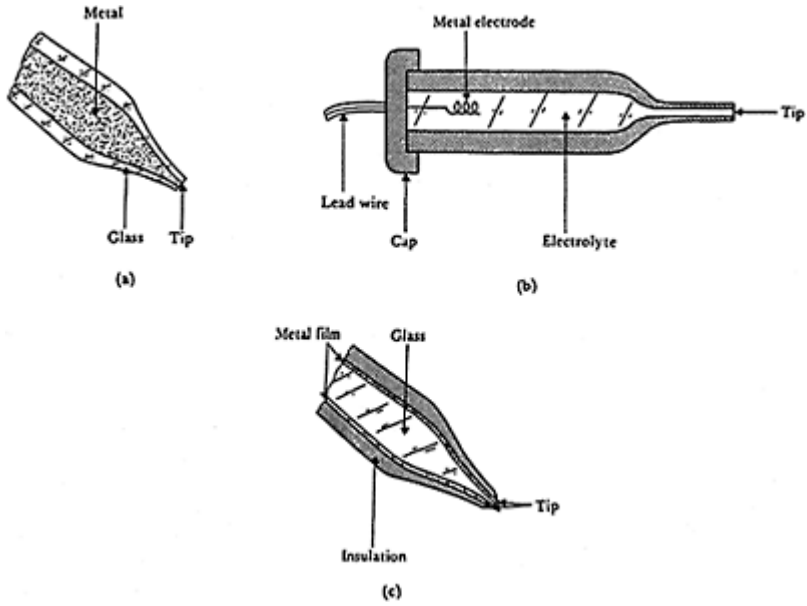


FIGURE 10.6 Microelectrodes: *(a)* metal, *(b)* micropipette, *(c)* thin metal film on micropipette. (Reprinted with permission from Webster JC (Ed). 1992. *Medical Instrumentation:*

Application and Design, Houghton Mifflin, Boston.)

can be used for neurophysiologic studies. Their principal disadvantage is the difficulty encountered in their fabrication and their high source impedance.

The second and most frequently used type of microelectrode is the glass micropipette. This structure, as illustrated in Fig. 10.6*b* consists of a fine glass capillary drawn to a very narrow point and filled with an electrolytic solution. The point can be as narrow as a fraction of a micrometer, and the dimensions of this electrode are strongly dependent on the skill of the individual drawing the tip. The electrolytic solution in the lumen serves as the contact between the interior of the cell through which the tip has been impaled and a larger conventional electrode located in the shank of the pipette. These electrodes also suffer from high source impedances and fabrication difficulty.

A combined form of these two types of electrodes can be achieved by depositing a metal film over the outside surface of a glass micropipette as shown in Fig. 10.6*c*. In this case, the strength and smaller dimensions of the micropipette can be used to support films of various metals that are insulated by an additional film up to a point very close to the actual tip of the electrode structure. These electrodes have been manufactured in quantity and made available as commercial products. Since they combine the features of both the metal and the micropipette electrodes, they also suffer from many of the same limitations. They do, however, have the advantage of flexibility due to the capability of being able to make films of different metals on the micropipette surface without having to worry about the strength of the metal, as would be the case if the metal were used alone.

Electrodes Fabricated Using Microelectronic Technology

Modern microelectronic technology can be used to fabricate many different types of electrodes for specific biomedical applications. For example, dry electrodes with high source resistances or microelectrodes with similar characteristics can be improved by incorporating a microelectronic amplifier for impedance conversion right on the electrode itself. In the case of the conventional-sized electrodes, a metal disk 5 to 10 mm in diameter can have a high input impedance microelectronic amplifier configured as a follower integrated into the back of the electrode so that localized processing of the high source impedance signal can produce one of lower, more practical impedance for signal transmission [8]. Single- and multiple-element electrodes can be made from thin-film or silicon technology. Mastrototaro and colleagues have demonstrated probes for measuring intramyocardial potentials using thin, patterned gold films on polyimide or oxidized molybdenum substrates [9]. When electrodes are made from pieces of micromachined silicon, it is possible to integrate an amplifier directly into the electrode [10]. Multichannel amplifiers or multiplexers can be used with multiple electrodes on the same probe. Electrodes for contact with individual nerve fibers can be fabricated using micromachined holes in a silicon chip that are just big enough to pass a single growing axon. Electrical contacts on the sides of these holes can then be used to pick up electrical activity from these nerves [11]. These examples are just a few of the many possibilities that can be realized using microelectronics and three-dimensional micromachining technology to fabricate specialized electrodes.

10.4 Biomedical Applications

Electrodes can be used to perform a wide variety of measurements of bioelectric signals. An extensive review of this would be beyond the scope of this chapter, but some typical examples of applications are highlighted in Table 10.4. The most popular application for biopotential electrodes is in obtaining the electrocardiogram for diagnostic and patient-monitoring applications. A substantial commercial market exists for various types of electrocardiographic electrodes, and many of the forms described in the previous section are available commercially. Other electrodes for measuring bioelectric potentials for application in diagnostic medicine are indicated in Table 10.4. Research applications of biopotential electrodes are highly varied and specific for individual studies. Although a few examples are given in Table 10.4, the field is far too broad to be completely covered here.

Biopotential electrodes are one of the most common biomedical sensors used in clinical medicine. Although their basic principle of operation is the same for most applications, they take on many forms

TABLE 10.4 Examples of Applications of Biopotential Electrodes

Application	Biopotential	Type of Electrode
Cardiac monitoring	ECG	Ag/AgCl with sponge
		Ag/AgCl with hydrogel
Infant cardiopulmonary monitoring	ECG impedance	Ag/AgCl with sponge
		Ag/AgCl with hydrogel
		Thin-film
		Filled elastomer dry
Sleep encephalography	EEG	Gold cups
		Ag/AgCl cups
		Active electrodes
Diagnostic muscle activity	EMG	Needle
Cardiac electrograms	Electrogram	Intracardiac probe
Implanted telemetry of biopotentials	ECG	Stainless steel wire loops
	EMG	Platinum disks
Eye movement	EOG	Ag/AgCl with hydrogel

and are used in the measurement of many types of bioelectric phenomena. They will continue to play an important role in biomedical instrumentation systems.

References

1. Yoshida T, Hayashi K, Toko K. 1988. The effect of anoxia on the spatial pattern of electric potential formed along the root. *Ann Bot* 62(5):497.
2. Plonsey R, Barr RC. 1988. *Bioelecricity,* New York, Plenum.
3. Weast RC (ed). 1974. *Handbook of Chemistry and Physics,* 55th ed, Boca Raton, Fla, CRC Press.
4. Webster JG (ed). 1992. *Medical Instrumentation: Application and Design,* Boston, Houghton Mifflin.
5. Janz GI, Ives DJG. 1968. Silver—Silver chloride electrodes. *Ann N Y Acad Sci* 148:210.
6. Geddes LA. 1972. *Electrodes and the Measurement of Bioelectric Events,* New York, Wiley.
7. Caldwell CW, Reswick JB. 1975. A percutaneous wire electrode for chronic research use. *IEEE Trans Biomed Eng* 22:429.
8. Ko WH, Hynecek J. 1974. Dry electrodes and electrode amplifiers. In HA Miller, Harrison DC (Eds), *Biomedical Electrode Technology,* pp 169–181, New York, Academic Press.
9. Mastrototaro JJ, Massoud HZ, Pilkington TC, et al, 1992. Rigid and flexible thin-film microelectrode arrays for transmural cardiac recording. *IEEE Trans Biomed Eng* 39(3):271.
10. Wise KD, Najafi K, Ji J, et al. 1990. Micromachined silicon microprobes for CNS recording and stimulation. *Proc Ann Conf IEEE Eng Med Biol Soc* 12:2334.
11. Edell DJ. 1986. A peripheral nerve information transducer for amputees: Long-term multichannel recordings from rabbit peripheral nerves. *IEEE Trans Biomed Eng* 33:203.

Further Information

Good overviews of biopotential electrodes are found in Geddes LA, 1972, *Electrodes and the Measurement of Bioelectric Events,* New York, Wiley; and Ferris CD, 1974, *Introduction to Bioelectrodes,* New York, Plenum. Even though these references are more than 20 years old, they clearly cover the field, and little has changed since these books were written.

Overviews of biopotential electrodes are found in chapters of two works edited by John Webster. Chapter 5 of his textbook, *Medical Instrumentation: Application and Design,* covers the material of this chapter in more detail, and there is a section on "Bioelectrodes" in his *Encyclopedia on Medical Devices and Instrumentation,* published by Wiley in 1988.

The journals, *IEEE Transactions on Biomedical Engineering* and *Medical and Biological Engineering and Computing,* are good sources of recent research on biopotential electrodes.

11
Electrochemical Sensors

Chung-Chiun Liu
Electronics Design Center and
Edison Sensor Technology Center,
Case Western Reserve University

Electrochemical sensors have been used extensively either as a whole or an integral part of a chemical and biomedical sensing element. For instance, blood gas (PO_2, PCO_2, and pH) sensing can be accomplished entirely by electrochemical means. Many important biomedical enzymatic sensors, including glucose sensors, incorporate an enzymatic catalyst and an electrochemical sensing element. The Clark type of oxygen sensor [Clark, 1956] is a well-known practical biomedical sensor based on electrochemical principles, an amperometric device. Electrochemical sensors generally can be categorized as conductivity/ capacitance, potentiometric, amperometric, and voltammetric sensors. The amperometric and voltammetric sensors are characterized by their current-potential relationship with the electrochemical system and are less well-defined. Amperometric sensors can also be viewed as a subclass of voltammetric sensors.

Electrochemical sensors are essentially an electrochemical cell that employs a two- or three-electrode arrangement. Electrochemical sensor measurement can be made at steady state or transient. The applied current or potential for electrochemical sensors may vary according to the mode of operation, and the selection of the mode is often intended to enhance the sensitivity and selectivity of a particular sensor. The general principles of electrochemical sensors have been extensively discussed in many electroanalytic references. However, many electroanalytic methods are not practical in biomedical sensing applications. For instance, dropping mercury electrode polarography is a well-established electroanalytic method, yet its usefulness in biomedical sensor development, particularly for potential in vivo sensing, is rather limited. In this chapter, we shall focus on the electrochemical methodologies that are useful in biomedical sensor development.

11.1 Conductivity/Capacitance Electrochemical Sensors

Measurement of the conductivity of an electrochemical cell can be the basis for an electrochemical sensor. This differs from an electrical (physical) measurement, for the electrochemical sensor measures the conductivity change of the system in the presence of a given solute concentration. This solute is often the sensing species of interest. Electrochemical sensors may also involve a capacitative impedance resulting from the polarization of the electrodes and the faradaic or charge transfer processes.

0-8493-1813-0/03/$0.00+$ 1.50

It has been established that the conductance of a homogeneous solution is directly proportional to the cross-sectional area perpendicular to the electrical field and inversely proportional to the segment of solution along the electrical field. Thus, the conductance of this solution (electrolyte), G(Ω^{-1}), can be expressed as

$$G = \sigma A / L , \tag{11.1}$$

where A is the cross-sectional area (in cm^2), L is the segment of the solution along the electrical field (in cm), and σ (in Ω^{-1} cm^{-1}) is the specific conductivity of the electrolyte and is related quantitatively to the concentration and the magnitude of the charges of the ionic species. For a practical conductivity sensor, A is the surface of the electrode, and L is the distance between the two electrodes.

Equivalent and molar conductivities are commonly used to express the conductivity of the electrolyte. Equivalent conductance depends on the concentration of the solution. If the solution is a strong electrolyte, it will completely dissociate the components in the solution to ionic forms. Kohlrauch [MacInnes, 1939] found that the equivalent conductance of a strong electrolyte was proportional to the square root of its concentration. However, if the solution is a weak electrolyte that does not completely dissociate the components in the solution to respective ions, the above observation by Kohlrauch is not applicable.

The formation of ions leads to consideration of their contribution to the overall conductance of the electrolyte. The equivalent conductance of a strong electrolyte approaches a constant limiting value at infinite dilution, namely,

$$\Lambda_o = \Lambda_{\lim \to o} = \lambda_o^+ + \lambda_o^- , \tag{11.2}$$

where Λ_o is the equivalent conductance of the electrolyte at infinite dilution and λ^+_o and λ^-_o are the ionic equivalent conductance of cations and anions at infinite dilution, respectively.

Kohlrauch also established the law of independent mobilities of ions at infinite dilution. This implies that Λ_o at infinite dilution is a constant at a given temperature and will not be affected by the presence of other ions in the electrolytes. This provides a practical estimation of the value of Λ_o from the values of λ^+_o and λ^-_o. As mentioned, the conductance of an electrolyte is influenced by its concentration. Kohlrausch stated that the equivalent conductance of the electrolyte at any concentration C in mol/1 or any other convenient units can be expressed as

$$\Lambda = \Lambda_o - \beta C^{0.5} , \tag{11.3}$$

where β is a constant depending on the electrolyte.

In general, electrolytes can be classified as weak electrolytes, strong electrolytes, and ion-pair electrolytes. Weak electrolytes only dissociate to their component ions to a limited extent, and the degree of the dissociation is temperature dependent. However, strong electrolytes dissociate completely, and Eq. (11.3) is applicable to evaluate its equivalent conductance. Ion-pair electrolytes can by characterized by their tendency to form ion pairs. The dissociation of ion pairs is similar to that of a weak electrolyte and is

affected by ionic activities. The conductivity of ion-pair electrolytes is often nonlinear related to its concentration.

The electrolyte conductance measurement technique, in principle, is relatively straightforward. However, the conductivity measurement of an electrolyte is often complicated by the polarization of the electrodes at the operating potential. Faradaic or charge transfer processes occur at the electrode surface, complicating the conductance measurement of the system. Thus, if possible, the conductivity electrochemical sensor should operate at a potential where no faradaic processes occur. Also, another important consideration is the formation of the double layer adjacent to each electrode surface when a potential is imposed on the electrochemical sensor. The effect of the double layer complicates the interpretation of the conductivity measurement and is usually described by the Warburg impedance. Thus, even in the absence of faradaic processes, the potential effect of the double layer on the conductance of the electrolyte must be carefully assessed. The influence of a faradaic process can be minimized by maintaining a high center constant, *L/A*, of the electrochemical conductivity sensor, so that the cell resistance lies in the region of 1 to 50 kΩ. This implies the desirable feature of a small electrode surface area and a relatively large distance between the two electrodes. Yet, a large electrode surface area enhances the accuracy of the measurement, since a large deviation from the null point facilitates the balance of the Wheatstone bridge, resulting in improvement of sensor sensitivity. These opposing features can be resolved by using a multiple-sensing electrode configuration in which the surface area of each electrode element is small compared to the distance between the electrodes. The multiple electrodes are connected in parallel, and the output of the sensor represents the total sum of the current through each pair of electrodes. In this mode of measurement, the effect of the double layer is included in the conductance measurement. The effects of both the double layers and the faradaic processes can be minimized by using a high-frequency, low-amplitude alternating current. The higher the frequency and the lower the amplitude of the imposed alternating current, the closer the measured value is to the true conductance of the electrolyte.

11.2 Potentiometric Sensors

When a redox reaction, Ox+Ze=Red, takes place at an electrode surface in an electrochemical cell, a potential may develop at the electrode-electrolyte interface. This potential may then be used to quantify the activity (on concentration) of the species involved in the reaction forming the fundamental of potentiometric sensors.

The above reduction reaction occurs at the surface of the cathode and is defined as a *half-cell reaction*. At thermodynamic equilibrium, the Nernst equation is applicable and can be expressed as:

$$E = E^{o} + \frac{RT}{ZF} \ln\left(\frac{a_{ox}}{a_{red}}\right), \tag{11.4}$$

where E and $E°$ are the measured electrode potential and the electrode potential at standard state, respectively; a_{ox} and a_{red} are the activities of Ox (reactant in this case) and

Red (product in this case), respectively; Z is the number of electrons transferred, F the Faraday constant, R the gas constant, and T the operating temperature in the absolute scale. In the electrochemical cell, two half-cell reactions will take place simultaneously. However, for sensing purposes, only one of the two half-cell reactions should involve the species of interest, and the other half-cell reaction is preferably reversible and noninterfering. As indicated in Eq. (11.4), a linear relation exists between the measured potential E and the natural logarithm of the ratio of the activities of the reactant and product. If the number of electrons transferred, Z, is one, at ambient temperature (25°C or 298°K) the slope is approximately 60 mV/decade. This slope value governs the sensitivity of the potentiometric sensor.

Potentiometric sensors can be classified based on whether the electrode is inert or active. An inert electrode does not participate in the half-cell reaction and merely provides the surface for the electron transfer or provides a catalytic surface for the reaction. However, an active electrode is either an ion donor or acceptor in the reaction. In general, there are three types of active electrodes: the metal/metal ion, the metal/insoluble salt or oxide, and metal/metal chelate electrodes.

Noble metals such as platinum and gold, graphite, and glassy carbon are commonly used as inert electrodes on which the half-cell reaction of interest takes place. To complete the circuitry for the potentiometric sensor, the other electrode is usually a reference electrode on which a noninterference half-cell reaction occurs. Silver-silver chloride and calomel electrodes are the most commonly used reference electrodes. Calomel consists of $Hg/HgCl_2$ and is less desirable for biomedical systems in terms of toxicity.

An active electrode may incorporate chemical or biocatalysts and is involved as either an ion donor or acceptor in the half-cell reaction. The other half-cell reaction takes place on the reference electrode and should also be noninterfering.

If more than a single type of ion contributes to the measured potential in Eq. (11.4), the potential can no longer be used to quantify the ions of interest. This is the interference in a potentiometric sensor. Thus, in many cases, the surface of the active electrode often incorporates a specific functional membrane which may be ion-selective, ion-permeable, or have ion-exchange properties. These membranes tend to selectivity permit the ions of interest to diffuse or migrate through. This minimizes the ionic interference.

Potentiometric sensors operate at thermodynamic equilibrium conditions. Thus, in practical potentiometric sensing, the potential measurement needs to be made under zero-current conditions. Consequently, a high-input impedance electrometer is often used for measurements. Also, the response time for a potentiometric sensor to reach equilibrium conditions in order to obtain a meaningful reading can be quite long. These considerations are essential in the design and selection of potentiometric sensors for biomedical applications.

11.3 Voltammetric Sensors

The current-potential relationship of an electrochemical cell provides the basis for voltammetric sensors. Amperometric sensors, that are also based on the current-potential relationship of the electrochemical cell, can be considered a subclass of voltammetric

sensors. In amperometric sensors, a fixed potential is applied to the electrochemical cell, and a corresponding current, due to a reduction or oxidation reaction, is then obtained. This current can be used to quantify the species involved in the reaction. The key consideration of an amperometric sensor is that it operates at a fixed potential. However, a voltammetric sensor can operate in other modes such as linear cyclic voltammetric modes. Consequently, the respective current potential response for each mode will be different.

In general, voltammetric sensors examine the concentration effect of the detecting species on the current-potential characteristics of the reduction or oxidation reaction involved.

The mass transfer rate of the detecting species in the reaction onto the electrode surface and the kinetics of the faradaic or charge transfer reaction at the electrode surface directly affect the current-potential characteristics. This mass transfer can be accomplished through *(a)* an ionic migration as a result of an electric potential gradient, *(b)* a diffusion under a chemical potential difference or concentration gradient, and *(c)* a bulk transfer by natural or forced convection. The electrode reaction kinetics and the mass transfer processes contribute to the rate of the faradaic process in an electrochemical cell. This provides the basis for the operation of the voltammetric sensor. However, assessment of the simultaneous mass transfer and kinetic mechanism is rather complicated. Thus, the system is usually operated under definitive hydrodynamic conditions. Various techniques to control either the potential or current are used to simplify the analysis of the voltammetric measurement. A description of these techniques and their corresponding mathematical analyses are well documented in many texts on electrochemistry or electroanalysis [Adams, 1969; Bard and Faulkner, 1980; Lingane, 1958; Macdonald, 1977; Murray and Reilley, 1966].

A preferred mass transfer condition is total diffusion, which can be described by Fick's law of diffusion. Under this condition, the cell current, a measure of the rate of the faradaic process at an electrode, usually increases with increases in the electrode potential. This current approaches a limiting value when the rate of the faradaic process at the electrode surface reaches its maximum mass transfer rate. Under this condition, the concentration of the detecting species at the electrode surface is considered as zero and is diffusional mass transfer. Consequently, the limiting current and the bulk concentration of the detecting species can be related by

$$i = ZFkmC^* \tag{11.5}$$

where *km* is the mass transfer coefficient and C^* is the bulk concentration of the detecting species. At the other extreme, when the electrode kinetics are slow compared with the mass transfer rate, the electrochemical system is operated in the reaction kinetic control regime. This usually corresponds to a small overpotential. The limiting current and the bulk concentration of the detecting species can be related as

$$i = ZFkcC^* \tag{11.6}$$

where *kc* is the kinetic rate constant for the electrode process. Both Eqs. (11.5) and (11.6) show the linearrelationship between the limiting current and the bulk concentration of the detecting species. In manycases, the current does not tend to a limiting value with

an increase in the electrode potential. This isbecause other faradaic or nonfaradaic processes become active, and the cell current represents thecumulative rates of all active electrode processes. The relative rates of these processes, expressing currentefficiency, depend on the current density of the electrode. Assessment of such a system is rather complicated, and the limiting current technique may become ineffective.

When a voltammetric sensor operates with a small overpotential, the rate of faradaic reaction is also small; consequently, a high-precision instrument for the measurement is needed. An amperometric sensor is usually operated under limiting current or relatively small overpotential conditions. Amperometric sensors operate under an imposed fixed electrode potential. Under this condition, the cell current can be correlated with the bulk concentration of the detecting species (the solute). This operating mode is commonly classified as amperometric in most sensor work, but it also is referred to as the *chronosuperometric* method, since time is involved.

Voltammetric sensors can be operated in a linear or cyclic sweep mode. Linear sweep voltammetry involves an increase in the imposed potential linearly at a constant scanning rate from an initial potential to a defined upper potential limit. This is the so-called potential window. The current-potential curve usually shows a peak at a potential where the oxidation or reduction reaction occurs. The height of the peak current can be used for the quantification of the concentration of the oxidation or reduction species. Cyclic voltammetry is similar to the linear sweep voltammetry except that the electrode potential returns to its initial value at a fixed scanning rate. The cyclic sweep normally generates the current peaks corresponding to the oxidation and reduction reactions. Under these circumstances, the peak current value can relate to the corresponding oxidation or reduction reaction. However, the voltammogram can be very complicated for a system involving adsorption (nonfaradaic processes) and charge processes (faradaic processes). The potential scanning rate, diffusivity of the reactant, and operating temperature are essential parameters for sensor operation, similar to the effects of these parameters for linear sweep voltammograms. The peak current may be used to quantify the concentration of the reactant of interest, provided that the effect of concentration on the diffusivity is negligible. The potential at which the peak current occurs can be used in some cases to identify the reaction, or the reactant. This identification is based on the half-cell potential of the electrochemical reactions, either oxidation or reduction. The values of these half-cell reactions are listed extensively in handbooks and references.

The described voltammetric and amperometric sensors can be used very effectively to carry out qualitative and quantitative analyses of chemical and biochemical species. The fundamentals of this sensing technique are well established, and the critical issue is the applicability of the technique to a complex, practical environment, such as in whole blood or other biologic fluids. This is also the exciting challenge of designing a biosensor using voltammetric and amperometric principles.

11.4 Reference Electrodes

Potentiometric, voltammetric, and amperometric sensors employ a reference electrode. The reference electrode in the case of potentiometric and amperometric sensors serves as a counter electrode to complete the circuitry. In either case, the reaction of interest takes

place at the surface of the working electrode, and this reaction is either an oxidation or reduction reaction. Consequently, the reaction at the counter electrode, i.e., the reference electrode, is a separate reduction or oxidation reaction, respectively. It is necessary that the reaction occurring at the reference electrode does not interfere with the reaction at the working electrode. For practical applications, the reaction occurring at the reference electrode should be highly reversible and, as stated, does not contribute to the reaction at the working electrode. In electrochemistry, the hydrogen electrode is universally accepted as the primary standard with which other electrodes are compared. Consequently, the hydrogen electrode serves extensively as a standard reference. A hydrogen reference electrode is relatively simple to prepare. However, for practical applications hydrogen reference electrodes are too cumbersome to be useful in practice.

A class of electrode called the *electrode of the second kind*, which forms from a metal and its sparingly soluble metal salt, finds use as the reference electrode. The most common electrode of this type includes the calomel electrode, $Hg/HgCl_2$ and the silver-silver chloride electrode, Ag/AgCl. In biomedical applications, particularly in *in vivo* applications, Ag/AgCl is more suitable as a reference electrode.

An Ag/AgCl electrode can be small, compact, and relatively simple to fabricate. As a reference electrode, the stability and reproducibility of an Ag/AgCl electrode is very important. Contributing factors to instability and poor reproducibility of Ag/AgCl electrodes include the purity of the materials used, the aging effect of the electrode, the light effect, and so on. When in use, the electrode and the electrolyte interface contribute to the stability of the reference electrode. It is necessary that a sufficient quantity of Cl^- ions exists in the electrolyte when the Ag/AgCl electrode serves as a reference. Therefore, other silver-silver halides such as Ag/AgBr or Ag/AgI electrodes are used in cases where these other halide ions are present in the electrolyte.

In a voltammetric sensor, the reference electrode serves as a true reference for the working electrode, and no current flows between the working and reference electrodes. Nevertheless, the stability of the reference electrode remains essential for a voltammetric sensor.

11.5 Summary

Electrochemical sensors are used extensively in many biomedical applications including blood chemistry sensors, PO_2, PCO_2, and pH electrodes. Many practical enzymatic sensors, including glucose and lactate sensors, also employ electrochemical sensors as sensing elements. Electrochemically based biomedical sensors have found *in vivo* and *in vitro* applications. We believe that electrochemical sensors will continue to be an important aspect of biomedical sensor development.

References

Adams RN. 1969. *Electrochemistry at Solid Electrodes*, New York, Marcel Dekker.

Bard A, Faulkner LR. 1980. *Electrochemical Methods*, New York, Wiley.

Clark LC Jr. 1956. Monitor and control of blood and tissue oxygen tissues. *Trans. Am. Soc. Artif. Organs* 2:41.

Lingane JJ. 1958. *Electroanalytical Chemistry*, New York, London, Interscience.
Macdonald DD. 1977. *Transient Techniques in Electrochemistry*, New York, Plenum.
MacInnes DA. 1939. *The Principles of Electrochemistry*, New York, Reinhold.
Murray RW, Reilley CN. 1996. *Electroanalytical Principles*, New York-London, Interscience.

12
Optical Sensors

Yitzhak Mendelson
Worcester Polytechnic Institute

Optical methods are among the oldest and best-established techniques for sensing biochemical analytes. Instrumentation for optical measurements generally consists of a light source, a number of optical components to generate a light beam with specific characteristics and to direct this light to some modulating agent, and a photodetector for processing the optical signal. The central part of an optical sensor is the modulating component, and a major part of this chapter will focus on how to exploit the interaction of an analyte with optical radiation in order to obtain essential biochemical information.

The number of publications in the field of optical sensors for biomedical applications has grown significantly during the past two decades. Numerous scientific reviews and historical perspectives have been published, and the reader interested in this rapidly growing field is advised to consult these sources for additional details. This chapter will emphasize the basic concept of typical optical sensors intended for continuous *in vivo* monitoring of biochemical variables, concentrating on those sensors that have generally progressed beyond the initial feasibility stage and reached the promising stage of practical development or commercialization.

Optical sensors are usually based on optical fibers or on planar waveguides. Generally, there are three distinctive methods for quantitative optical sensing at surfaces:

1. The analyte directly affects the optical properties of a waveguide, such as evanescent waves (electromagnetic waves generated in the medium outside the optical waveguide when light is reflected from within) or surface plasmons (resonances induced by an evanescent wave in a thin film deposited on a waveguide surface).
2. An optical fiber is used as a plain transducer to guide light to a remote sample and return light from the sample to the detection system. Changes in the intrinsic optical properties of the medium itself are sensed by an external spectrophotometer.
3. An indicator or chemical reagent placed inside, or on, a polymeric support near the tip of the optical fiber is used as a mediator to produce an observable optical signal. Typically, conventional techniques, such as absorption spectroscopy and fluorimetry, are employed to measure changes in the optical signal.

0-8493-1813-0/03/$0.00+$1.50

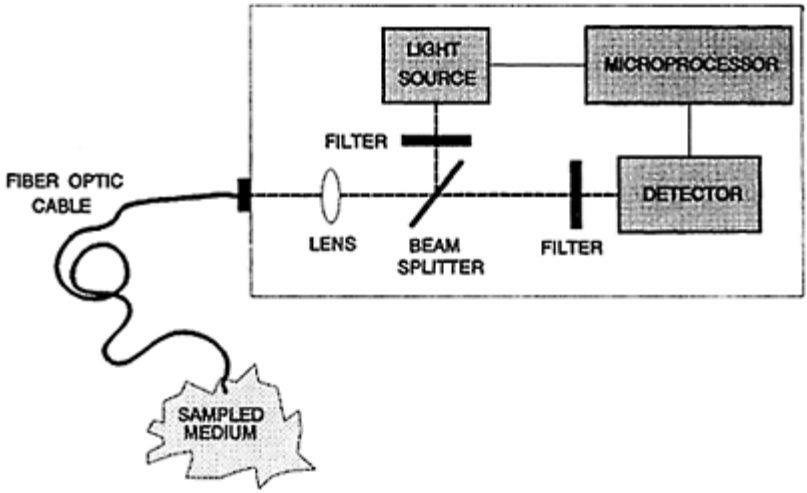


FIGURE 12.1 General diagram representing the basic building blocks of an optical instrument for optical sensor applications.

12.1 Instrumentation

The actual implementation of instrumentation designed to interface with optical sensors will vary greatly depending on the type of optical sensor used and its intended application. A block diagram of a generic instrument is illustrated in Fig. 12.1. The basic building blocks of such an instrument are the light source, various optical elements, and photodetectors.

Light Source

A wide selection of light sources are available for optical sensor applications. These include: highly coherent gas and semiconductor diode lasers, broad spectral band incandescent lamps, and narrow-band, solid-state, light-emitting diodes (LEDs). The important requirement of a light source is obviously good stability. In certain applications, for example in portable instrumentation, LEDs have significant advantages over other light sources because they are small and inexpensive, consume lower power, produce selective wavelengths, and are easy to work with. In contrast, tungsten lamps provide a broader range of wavelengths, higher intensity, and better stability but require a sizable power supply and can cause heating problems inside the apparatus.

Optical Elements

Various optical elements are used routinely to manipulate light in optical instrumentation. These include lenses, mirrors, light choppers, beam splitters, and couplers for directing the light from the light source into the small aperture of a fiber-optic sensor or a specific area on a waveguide surface and collecting the light from the sensor before it is processed by the photodetector. For wavelength selection, optical filters, prisms, and diffraction gratings are the most common components used to provide a narrow bandwidth of excitation when a broadwidth light source is utilized.

Photodetectors

In choosing photodetectors for optical sensors, a number of factors must be considered. These include sensitivity, detectivity, noise, spectral response, and response time. Photomultipliers and semiconductor quantum photodetectors, such as photoconductors and photodiodes, are both suitable. The choice, however, is somewhat dependent on the wavelength region of interest. Generally, both give adequate performance. Photodiodes are usually more attractive because of the compactness and simplicity of the circuitry involved.

Typically, two photodetectors are used in optical instrumentation because it is often necessary to include a separate reference detector to track fluctuations in source intensity and temperature. By taking a ratio between the two detector readings, whereby a part of the light that is not affected by the measurement variable is used for correcting any optical variations in the measurement system, a more accurate and stable measurement can be obtained.

Signal Processing

Typically, the signal obtained from a photodetector provides a voltage or a current proportional to the measured light intensity. Therefore, either simple analog computing circuitry (e.g., a current-to-voltage converter) or direct connection to a programmable gain voltage stage is appropriate. Usually, the output from a photodetector is connected directly to a preamplifier before it is applied to sampling and analog-to-digital conversion circuitry residing inside a computer.

Quite often two different wavelengths of light are utilized to perform a specific measurement. One wavelength is usually sensitive to changes in the species being measured, and the other wavelength is unaffected by changes in the analyte concentration. In this manner, the unaffected wavelength is used as a reference to compensate for fluctuation in instrumentation over time. In other applications, additional discriminations, such as pulse excitation or electronic background subtraction utilizing synchronized lock-in amplifier detection, are useful, allowing improved selectivity and enhanced signal-to-noise ratio.

12.2 Optical Fibers

Several types of biomedical measurements can be made by using either plain optical fibers as a remote device for detecting changes in the spectral properties of tissue and blood or optical fibers tightly coupled to various indicator-mediated transducers. The measurement relies either on direct illumination of a sample through the endface of the fiber or by excitation of a coating on the side wall surface through evanescent wave coupling. In both cases, sensing takes place in a region outside the optical fiber itself. Light emanating from the fiber end is scattered or fluoresced back into the fiber, allowing measurement of the returning light as an indication of the optical absorption or fluorescence of the sample at the fiber optic tip.

Optical fibers are based on the principle of total internal reflection. Incident light is transmitted through the fiber if it strikes the cladding at an angle greater than the so-called critical angle, so that it is totally internally reflected at the core/cladding interface. A typical instrument for performing fiber-optic sensing consists of a light source, an optical coupling arrangement, the fiber-optic light guide with or without the necessary sensing medium incorporated at the distal tip, and a light detector.

A variety of high-quality optical fibers are available commercially for biomedical sensor applications, depending on the analytic wavelength desired. These include plastic, glass, and quartz fibers that cover the optical spectrum from the UV through the visible to the near IR region. On one hand, plastic optical fibers have a larger aperture and are strong, inexpensive, flexible, and easy to work with but have poor UV transmission below 400 nm. On the other hand, glass and quartz fibers have low attenuation and better transmission in the UV but have small apertures, are fragile, and present a potential risk in *in vivo* applications.

Probe Configurations

There are many different ways to implement fiber-optic sensors. Most fiber-optic chemical sensors employ either a single-fiber configuration, where light travels to and from the sensing tip in one fiber, or a double-fiber configuration, where separate optical fibers are used for illumination and detection. A single fiber-optic configuration offers the most compact and potentially least expensive implementation. However, additional challenges in instrumentation are involved in separating the illuminating signal from the composite signal returning for processing.

The design of intravascular catheters requires special considerations related to the sterility and bio-compatibility of the sensor. For example, intravascular fiber-optic sensors must be sterilizable and their material nonthrombogenic and resistant to platelet and protein deposition. Therefore, these catheters are typically made of materials covalently bound with heparin or antiplatelet agents. The catheter is normally introduced into the jugular vein via a peripheral cutdown and a slow heparin flush is maintained until it is removed from the blood.

Optical Fiber Sensors

Advantages cited for fiber sensors include their small size and low cost. In contrast to electrical measurements, where the difference of two absolute potentials must be measured, fiber optics are self-contained and do not require an external reference signal. Because the signal is optical, there is no electrical risk to the patient, and there is no direct interference from surrounding electric or magnetic fields. Chemical analysis can be performed in real time with almost an instantaneous response. Furthermore, versatile sensors can be developed that respond to multiple analytes by utilizing multiwavelength measurements.

Despite these advantages, optical fiber sensors exhibit several shortcomings. Sensors with immobilized dyes and other indicators have limited long-term stability, and their shelf life degrades over time. Moreover, ambient light can interfere with the optical measurement unless optical shielding or special time synchronous gating is performed.

Indicator-Mediated Transducers

Only a limited number of biochemical analytes have an intrinsic optical absorption that can be measured with sufficient selectivity directly by spectroscopic methods. Other species, particularly hydrogen, oxygen, carbon dioxide, and glucose, which are of primary interest in diagnostic applications, are not susceptible to direct photometry. Therefore, indicator-mediated sensors have been developed using specific reagents that are properly immobilized on the surface of an optical sensor.

The most difficult aspect of developing an optical biosensor is the coupling of light to the specific recognition element so that the sensor can respond selectively and reversibly to a change in the concentration of a particular analyte. In fiber-optic-based sensors, light travels efficiently to the end of the fiber where it exists and interacts with a specific chemical or biologic recognition element that is immobilized at the tip of the fiber optic. These transducers may include indicators and ionophores (i.e., ion-binding compounds) as well as a wide variety of selective polymeric materials. After the light interacts with the sample, the light returns through the same or a different optical fiber to a detector that correlates the degree of change with the analyte concentration.

Typical indicator-mediated fiber-optic-sensor configurations are shown schematically in Fig. 12.2. In *(a)* the indicator is immobilized directly on a membrane positioned at the end of a fiber. An indicator in the form of a powder can be either glued directly onto a membrane, as shown in *(b)*, or physically retained in position at the end of the fiber by a special permeable membrane *(c)*, a tubular capillary/ membrane *(d)*, or a hollow capillary tube *(e)*.

12.3 General Principles of Optical Sensing

Two major optical techniques are commonly available to sense optical changes at sensor interfaces. These are usually based on evanescent wave and surface plasmon resonance principles.

Evanescent Wave Spectroscopy

When light propagates along an optical fiber, it is not confined to the core region but penetrates to some extent into the surrounding cladding region. In this case, an electromagnetic component of the light

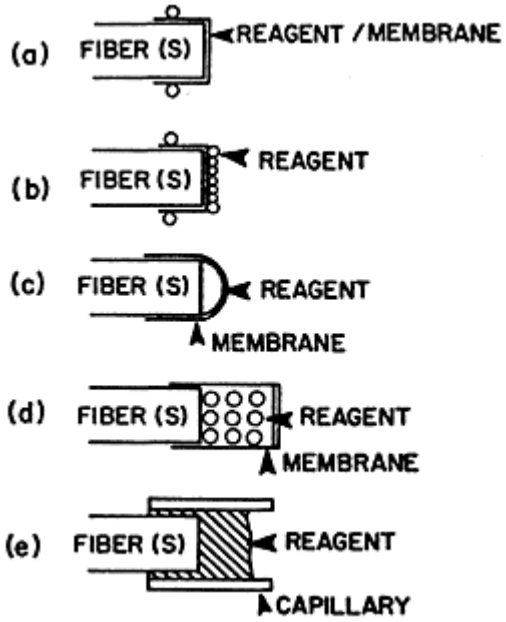


FIGURE 12.2 Typical configuration of different indicator-mediated fiber optic sensor tips (from Otto S.Wolfbeis, *Fiber Optic Chemical Sensors and Biosensors*, Vol. 1, CRC Press, Boca Raton, 1990).

penetrates a characteristic distance (on the order of one wavelength) beyond the reflecting surface into the less optically dense medium where it is attenuated exponentially according to Beer-Lambert's law (Fig. 12.3).

The evanescent wave depends on the angle of incidence and the incident wavelength. This phenomenon has been widely exploited to construct different types of optical sensors for biomedical applications. Because of the short penetration depth and the exponential decay of the intensity, the evanescent wave is absorbed mainly by absorbing compounds very close to the surface. In the case of particularly weak absorbing analytes, sensitivity can be enhanced by combining the evanescent wave principle with multiple internal reflections along the sides of an unclad portion of a fiber optic tip.

Instead of an absorbing species, a fluorophore can also be used. Light is absorbed by the fluorophore emitting detectable fluorescent light at a higher wavelength, thus

providing improved sensitivity. Evanescent wave sensors have been applied successfully to measure the fluorescence of indicators in solution, for pH measurement, and in immunodiagnostics.

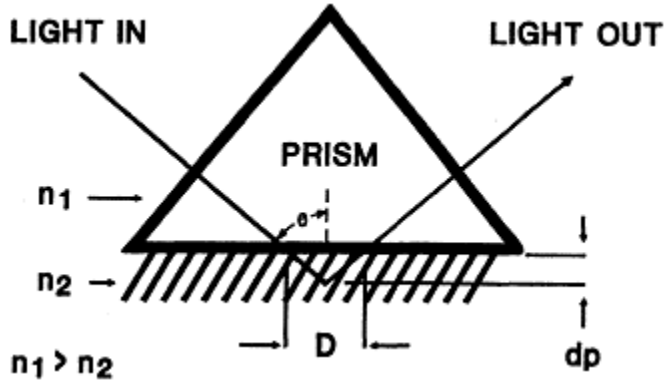


FIGURE 12.3 Schematic diagram of the path of a light ray at the interface of two different optical materials with index of refraction n_1 and n_2. The ray penetrates a fraction of a wavelength *(dp)* beyond the interface into the medium with the smaller refractive index.

Surface Plasmon Resonance

Instead of the dielectric/dielectric interface used in evanescent wave sensors, it is possible to arrange a dielectric/metal/dielectric sandwich layer such that when monochromatic polarized light (e.g., from a laser source) impinges on a transparent medium having a metallized (e.g., Ag or Au) surface, light is absorbed within the plasma formed by the conduction electrons of the metal. This results in a phenomenon known as *surface plasmon resonance* (SPR). When SPR is induced, the effect is observed as a minimum in the intensity of the light reflected off the metal surface.

As is the case with the evanescent wave, an SPR is exponentially decaying into solution with a penetration depth of about 20 nm. The resonance between the incident light and the plasma wave depends on the angle, wavelength, and polarization state of the incident light and the refractive indices of the metal film and the materials on either side of the metal film. A change in the dielectric constant or the refractive index at the surface causes the resonance angle to shift, thus providing a highly sensitive means of monitoring surface reactions.

The method of SPR is generally used for sensitive measurement of variations in the refractive index of the medium immediately surrounding the metal film. For example, if

an antibody is bound to or absorbed into the metal surface, a noticeable change in the resonance angle can be readily observed because of the change of the refraction index at the surface, assuming all other parameters are kept constant (Fig. 12.4). The advantage of this concept is the improved ability to detect the direct interaction between antibody and antigen as an interfacial measurement.

SPR has been used to analyze immunochemicals and to detect gases. The main limitation of SPR, however, is that the sensitivity depends on the optical thickness of the adsorbed layer, and, therefore, small molecules cannot be measured in very low concentrations.

12.4 Applications

12.4.1 Oximetry

Oximetry refers to the colorimetric measurement of the degree of oxygen saturation, that is, the relative amount of oxygen carried by the hemoglobin in the erythrocytes, by recording the variation in the color of deoxyhemoglobin (Hb) and oxyhemoglobin (HbO_2). A quantitative method for measuring blood oxygenation is of great importance in assessing the circulatory and respiratory status of a patient.

Various optical methods for measuring the oxygen saturation of arterial (SaO_2) and mixed venous (SvO_2) blood have been developed, all based on light transmission through, or reflecting from, tissue and blood. The measurement is performed at two specific wavelengths: λ_1, where there is a large difference in light absorbance between Hb and HbO_2 (e.g., 660 nm red light), and λ_2, which can be an isobestic wavelength (e.g., 805 nm infrared light), where the absorbance of light is independent of blood oxygenation, or a different wavelength in the infrared region (>805 nm), where the absorbance of Hb is slightly smaller than that of HbO_2.

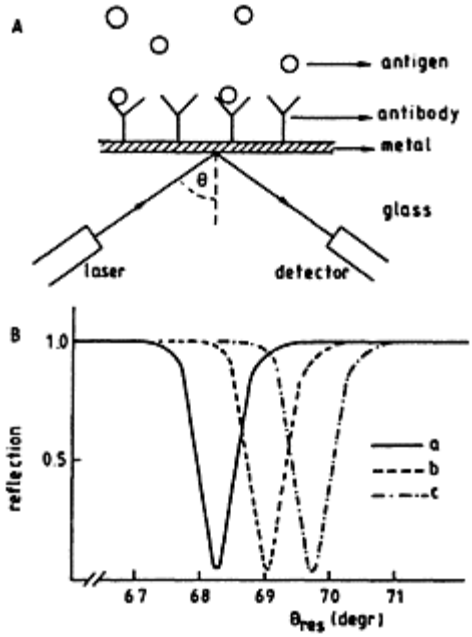


FIGURE 12.4 Surface plasmon resonance at the interface between a thin metallic surface and a liquid (A). A sharp decrease in the reflected light intensity can be observed in (B). The location of the resonance angle is dependent on the refractive index of the material present at the interface.

Assuming for simplicity that a hemolyzed blood sample consists of a two-component homogeneous mixture of Hb and HbO_2, and that light absorbance by the mixture of these two components is additive, a simple quantitative relationship can be derived for computing the oxygen saturation of blood:

$$\text{Oxygen saturation} = A - \mathrm{B}\left[\frac{\mathrm{OD}(\lambda_1)}{\mathrm{OD}(\lambda_2)}\right],$$

where *A* and *B* are coefficients that are functions of the specific absorptivities of Hb and HbO_2, and *OD* is the corresponding absorbance (optical density) of the blood.

Since the original discovery of this phenomenon more than 50 years ago, there has been progressive development in instrumentation to measure oxygen saturation along three

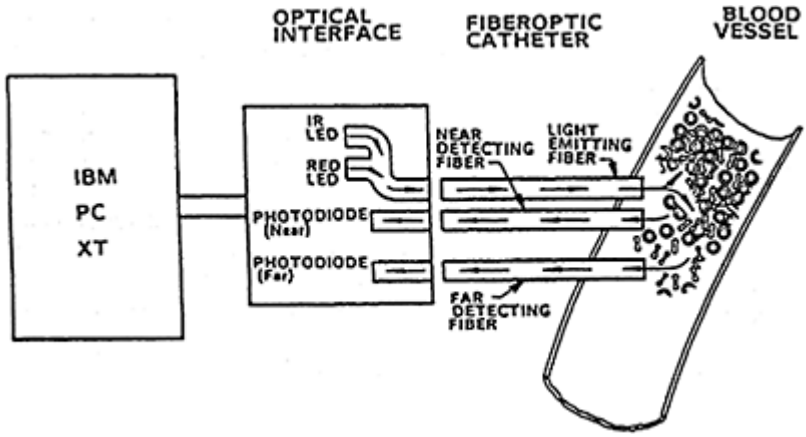


FIGURE 12.5 Principle of a three-fiber optical catheter for SvO_2/HCT measurement [2].

different paths: bench-top oximeters for clinical laboratories, fiber-optic catheters for invasive intravascular monitoring, and transcutaneous sensors, which are noninvasive devices placed against the skin.

Intravascular Fiber Optic SvO_2 Catheters

In vivo fiberoptic oximeters were first described in the early 1960s by Polanyi and Heir [1]. They demonstrated that in a highly scattering medium such as blood, where a very short path length is required for a transmittance measurement, a reflectance measurement was practical. Accordingly, they showed that a linear relationship exists between oxygen saturation and the ratio of the infrared-to-red (IR/R) light backscattered from the blood,

$$\text{oxygen saturation} = \mathbf{a} - \mathbf{b}\left(\text{IR}/\text{R}\right),$$

where **a** and **b** are catheter-specific calibration coefficients.

Fiber-optic SvO_2 catheters consist of two separate optical fibers. One fiber is used for transmitting the light to the flowing blood, and a second fiber directs the backscattered light to a photodetector. In some commercial instruments (e.g., Oximetrix), automatic compensation for hematocrit is employed utilizing three, rather than two, infrared reference wavelengths. Bornzin and coworkers [2] and Mendelson and coworkers [3] described a 5-lumen, 7.5F thermodilution catheter that is comprised of three unequally spaced optical fibers, each fiber 250 μm in diameter, and provides continuous SvO_2 reading with automatic corrections for hematocrit variations (Fig. 12.5).

Intravenous fiber-optic catheters are utilized in monitoring SvO_2 in the pulmonary artery and can be used to indicate the effectiveness of the cardiopulmonary system during cardiac surgery and in the ICU. Several problems limit the wide clinical application of intravascular fiberoptic oximeters. These include the dependence of the individual red

and infrared backscattered light intensities and their ratio on hematocrit (especially for SvO_2 below 80%), blood flow, motion artifacts due to catheter tip "whipping" against the blood vessel wall, blood temperature, and pH.

Noninvasive Pulse Oximetry

Noninvasive monitoring of SaO_2 by pulse oximetry is a rapidly growing practice in many fields of clinical medicine [4]. The most important advantage of this technique is the capability to provide continuous, safe, and effective monitoring of blood oxygenation at the patient's bedside without the need to calibrate the instrument before each use.

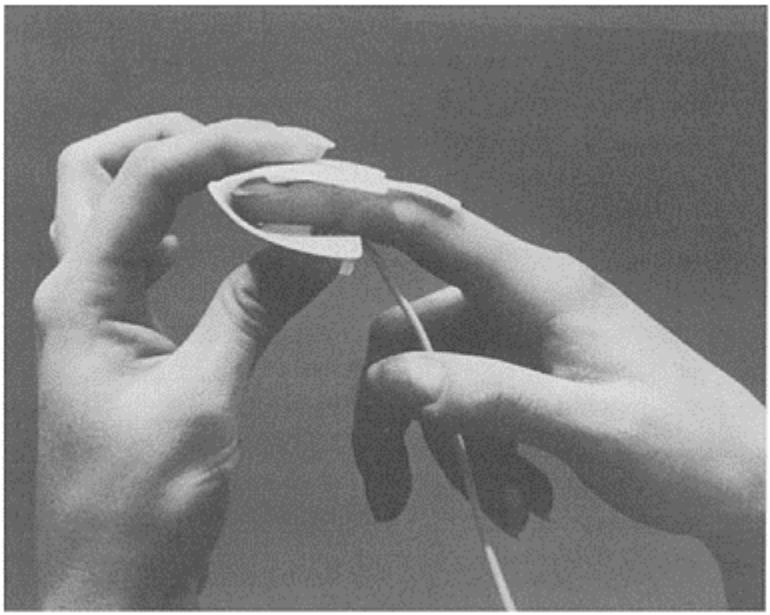


FIGURE 12.6 Disposable finger probe of a noninvasive pulse oximeter.

Pulse oximetry, which was first suggested by Aoyagi and colleagues [5] and Yoshiya and colleagues [6], relies on the detection of the time-variant photoplethysmographic signal, caused by changes in arterial blood volume associated with cardiac contraction. SaO_2 is derived by analyzing only the time-variant changes in absorbance caused by the pulsating arterial blood at the same red and infrared wavelengths used in conventional invasive type oximeters. A normalization process is commonly performed by which the pulsatile *(ac)* component at each wavelength, which results from the expansion and relaxation of the arterial bed, is divided by the corresponding nonpulsatile *(dc)* component of the photoplethysmogram, which is composed of the light absorbed by the bloodless tissue and the nonpulsatile portion of the blood compartment. This effective scaling process results in a normalized red/infrared ratio that is dependent on SaO_2 but is largely

independent of the incident light intensity, skin pigmentation, skin thickness, and tissue vasculature.

Pulse oximeter sensors consist of a pair of small and inexpensive red and infrared LEDs and a single, highly sensitive, silicon photodetector. These components are mounted inside a reusable rigid spring-loaded clip, a flexible probe, or a disposable adhesive wrap (Fig. 12.6). The majority of the commercially available sensors are of the transmittance type in which the pulsatile arterial bed, e.g., ear lobe, fingertip, or toe, is positioned between the LEDs and the photodetector. Other probes are available for reflectance (backscatter) measurement where both the LEDs and photodetectors are mounted side-by-side facing the skin [7,8].

Noninvasive Cerebral Oximetry

Another substance whose optical absorption in the near-infrared changes corresponding to its reduced and oxidized state is cytochrome aa3, the terminal member of the respiratory chain. Although the concentration of cytochrome aa3 is considerably lower than that of hemoglobin, advanced instrumentation including time-resolved spectroscopy and differential measurements is being used successfully to obtain noninvasive measurements of hemoglobin saturation and cytochrome aa3 by transilluminating areas of the neonatal brain [9–11].

Blood Gases

Frequent measurement of blood gases, i.e., oxygen partial pressure (PO_2), carbon dioxide partial pressure (PCO_2), and pH, is essential to clinical diagnosis and management of respiratory and metabolic problems in the operating room and the ICU. Considerable effort has been devoted over the last two decades to developing disposable extracorporeal and in particular intravascular fiber-optic sensors that can be used to provide continuous information on the acid-base status of a patient.

In the early 1970s, Lübbers and Opitz [12] originated what they called *optodes* (from the Greek, *optical path*) for measurements of important physiologic gases in fluids and in gases. The principle upon which these sensors was designed was a closed cell containing a fluorescent indicator in solution, with a membrane permeable to the analyte of interest (either ions or gases) constituting one of the cell walls. The cell was coupled by optical fibers to a system that measured the fluorescence in the cell. The cell solution would equilibrate with the PO_2 or PCO_2 of the medium placed against it, and the fluorescence of an indicator reagent in the solution would correspond to the partial pressure of the measured gas.

Extracorporeal Measurement

Following the initial feasibility studies of Lübbers and Opitz, Cardiovascular Devices (GDI, USA) developed a GasStat™ extracorporeal system suitable for continuous online monitoring of blood gases *ex vivo* during cardiopulmonary bypass operations. The system consists of a disposable plastic sensor connected inline with a blood loop through a fiber-optic cable. Permeable membranes separate the flowing blood from the system chemistry.

The CO_2-sensitive indicator consists of a fine emulsion of a bicarbonate buffer in a two-component silicone. The pH-sensitive indicator is a cellulose material to which hydroxypyrene trisulfonate (HPTS) is bonded covalently. The O_2-sensitive chemistry is composed of a solution of oxygen-quenching decacyclene in a one-component silicone covered with a thin layer of black PTFE for optical isolation and to render the measurement insensitive to the halothane anesthetic.

The extracorporeal device has two channels, one for arterial blood and the other for venous blood, and is capable of recording the temperature of the blood for correcting the measurements to 37°C. Several studies have been conducted comparing the specifications of the GasStat™ with that of intermittent blood samples analyzed on bench-top blood gas analyzers [13–15].

Intravascular Catheters

During the past decade, numerous efforts have been made to develop integrated fiber-optic sensors for intravascular monitoring of blood gases. A few commercial systems for monitoring blood gases and pH are currently undergoing extensive clinical testing. Recent literature reports of sensor performance show that considerable progress has been made mainly in improving the accuracy and reliability of these intravascular blood gas sensors [16–19].

Most fiber-optic intravascular blood gas sensors employ either a single- or a double-fiber configuration. Typically, the matrix containing the indicator is attached to the end of the optical fiber as illustrated in Fig. 12.7. Since the solubility of O_2 and CO_2 gases, as well as the optical properties of the sensing chemistry itself, are affected by temperature variations, fiber-optic intravascular sensors include a thermocouple or thermistor wire running alongside the fiber-optic cable to monitor and correct for temperature fluctuations near the sensor tip. A nonlinear response is characteristic of most chemical indicator sensors, so they are designed to match the concentration region of the intended application. Also, the response time of the optode is somewhat slower compared to electrochemical sensors.

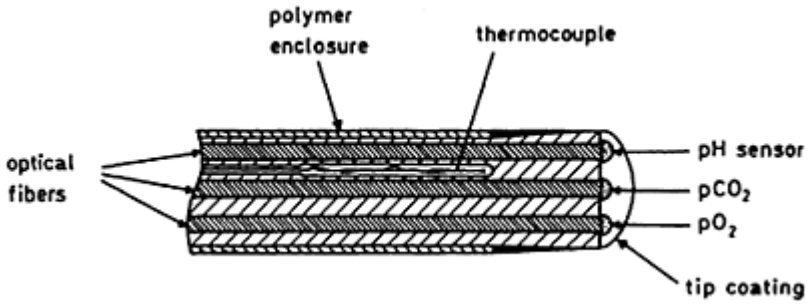


FIGURE 12.7 Principle diagram of an integrated fiber-optic blood gas catheter (from Otto S.Wolfbeis, *Fiber Optic Chemical Sensors and*

Biosensors, Vol. 2, CRC Press, Boca Raton, 1990).

Intravascular fiber-optic blood gas sensors are normally placed inside a standard 20-gauge catheter, which is sufficiently small to allow adequate spacing between the sensor and the catheter wall. The resulting lumen is large enough to permit the withdrawal of blood samples, introduction of a continuous heparin flush, and the recording of a blood pressure waveform. In addition, the optical fibers are encased in a protective tubing to contain any fiber fragments in case they break off.

pH Sensors

In 1976, Peterson and coworkers [20] originated the development of the first fiber-optic chemical sensor for physiological pH measurement. The basic idea was to contain a reversible, color-changing indicator at the end of a pair of optical fibers. The indicator, phenol red, was covalently bound to a hydrophilic polymer in the form of water-permeable microbeads. This technique stabilized the indicator concentration. The indicator beads were contained in a sealed hydrogen-ion-permeable envelope made out of a hollow cellulose tubing. In effect, this formed a miniature spectrophotometric cell at the end of the fibers and represented an early prototype of a fiber-optic chemical sensor.

The phenol red dye indicator is a weak organic acid, and the acid form (un-ionized) and base form (ionized) are present in a concentration ratio determined by the ionization constant of the acid and the pH of the medium according to the familiar Henderson-Hasselbalch equation. The two forms of the dye have different optical absorption spectra, so the relative concentration of one of the forms, which varies as a function of pH, can be measured optically and related to variations in pH. In the pH sensor, green (560 nm) and red (longer than 600 nm) light emerging from the end of one fiber passes through the dye and is reflected back into the other fiber by light-scatter ing particles. The green light is absorbed by the base form of the indicator. The red light is not absorbed by the indicator and is used as an optical reference. The ratio of green to red light is measured and is related to pH by an S-shaped curve with an approximate high-sensitivity linear region where the equilibrium constant *(pK)* of the indicator matches the pH of the solution.

The same principle can also be used with a reversible fluorescent indicator, in which case the concentration of one of the indicator forms is measured by its fluorescence rather than absorbance intensity. Light in the blue or UV wavelength region excites the fluorescent dye to emit longer wavelength light, and the two forms of the dye may have different excitation or emission spectra to allow their distinction.

The original instrument design for a pH measurement was very simple and consisted of a tungsten lamp for fiber illumination, a rotating filter wheel to select the green and red light returning from the fiber-optic sensor, and signal processing instrumentation to give a pH output based on the green-to-red ratio. This system was capable of measuring pH in the physiologic range between 7.0 to 7.4 with an accuracy and precision of 0.01 pH units. The sensor was susceptible to ionic strength variation in the order of 0.01 pH unit per 11% change in ionic strength.

Further development of the pH probe for practical use was continued by Markle and colleagues [21]. They designed the fiber-optic probe in the form of a 25-gauge (0.5 mm

o.d.) hypodermic needle, with an ion-permeable side window, using 75-mm-diameter plastic optical fibers. The sensor had a 90% response time of 30s. With improved instrumentation and computerized signal processing and with a three-point calibration, the range was extended to ±3 pH units, and a precision of 0.001 pH units was achieved.

Several reports have appeared suggesting other dye indicator systems that can be used for fiber-optic pH sensing [22]. A classic problem with dye indicators is the sensitivity of their equilibrium constant to ionic strength. To circumvent this problem, Wolfbeis and Offenbacher [23] and Opitz and Lübbers [24] demonstrated a system in which a dual sensor arrangement can measure ionic strength and pH and simultaneously can correct the pH measurement for variations in ionic strength.

PCO_2 Sensors

The PCO_2 of a sample is typically determined by measuring changes in the pH of a bicarbonate solution that is isolated from the sample by a CO_2-permeable membrane but remains in equilibrium with the CO_2. The bicarbonate and CO_2, as carbonic acid, form a pH buffer system, and, by the Henderson-Hasselbalch equation, hydrogen ion concentration is proportional to the pCO_2 in the sample. This measurement is done with either a pH electrode or a dye indicator in solution.

Vurek [25] demonstrated that the same techniques can also be used with a fiber-optic sensor. In his design, one plastic fiber carries light to the transducer, which is made of a silicone rubber tubing about 0.6 mm in diameter and 1.0 mm long, filled with a phenol red solution in a 35-mM bicarbonate. Ambient PCO_2 controls the pH of the solution, which changes the optical absorption of the phenol red dye. The CO_2 permeates through the rubber to equilibrate with the indicator solution. A second optical fiber carries the transmitted signal to a photodetector for analysis. The design by Zhujun and Seitz [26] uses a PCO_2 sensor based on a pair of membranes separated from a bifurcated optical fiber by a cavity filled with bicarbonate buffer. The external membrane is made of silicone, and the internal membrane is HPTS immobilized on an ion-exchange membrane.

PO_2 Sensors

The development of an indicator system for fiber-optic PO_2 sensing is challenging because there are very few known ways to measure PO_2 optically. Although a color-changing indicator would have been desirable, the development of a sufficiently stable indicator has been difficult. The only principle applicable to fiber optics appears to be the quenching effect of oxygen on fluorescence.

Fluorescence quenching is a general property of aromatic molecules, dyes containing them, and some other substances. In brief, when light is absorbed by a molecule, the absorbed energy is held as an excited electronic state of the molecule. It is then lost by coupling to the mechanical movement of the molecule (heat), reradiated from the molecule in a mean time of about 10 ns (fluorescence), or converted into another excited state with much longer mean lifetime and then reradiated (phosphorescence). Quenching reduces the intensity of fluorescence and is related to the concentration of the quenching molecules, such as O_2.

A fiber-optic sensor for measuring PO_2 using the principle of fluorescence quenching was developed by Peterson and colleagues [27]. The dye is excited at around 470 nm (blue) and fluoresces at about 515 nm (green) with an intensity that depends on the PO_2. The optical information is derived from the ratio of green fluorescence to the blue excitation light, which serves as an internal reference signal. The system was chosen for visible light excitation, because plastic optical fibers block light transmission at wavelengths shorter than 450 nm, and glass fibers were not considered acceptable for biomedical use.

The sensor was similar in design to the pH probe continuing the basic idea of an indicator packing in a permeable container at the end of a pair of optical fibers. A dye perylene dibutyrate, absorbed on a macroreticular polystyrene adsorbent, is contained in a oxygen-permeable porous polystyrene envelope. The ratio of green to blue intensity was processed according to the Stren-Volmer equation:

$$\frac{I_0}{I} = 1 + K\ \mathrm{PO}_2,$$

where I and I_o are the fluorescence emission intensities in the presence and absence of a quencher, respectively, and I is the Stern-Volmer quenching coefficient. This provides a nearly linear readout of PO_2 over the range of 0 to 150 mmHg (0 to 20 kPa), with a precision of 1 mm Hg (0.13 kPa). The original sensor was 0.5 mm in diameter, but it can be made much smaller. Although its response time in a gas mixture is a fraction of a second, it is slower in an aqueous system, about 1.5 min for 90% response.

Wolfbeis and coworkers [28] designed a system for measuring the widely used halothane anesthetic which interferes with the measurement of oxygen. This dual-sensor combination had two semipermeable membranes (one of which blocked halothane) so that the probe could measure both oxygen and halothane simultaneously. The response time of their sensor, 15 to 20 s for halothane and 10 to 15 s for oxygen, is considered short enough to allow gas analysis in the breathing circuit. Potential applications of this device include the continuous monitoring of halothane in breathing circuits and in the blood.

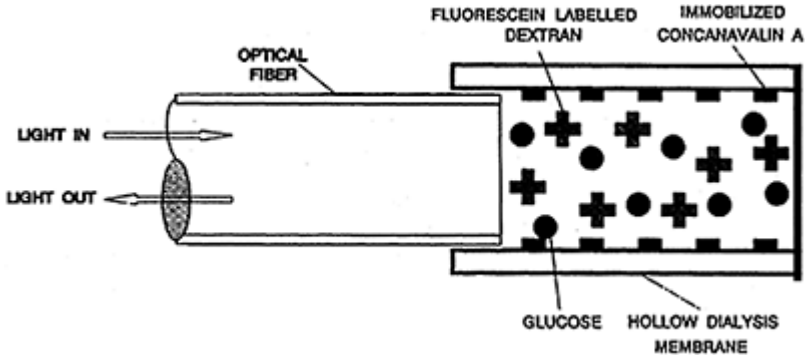


FIGURE 12.8 Schematic diagram of a competitive binding fluorescence

affinity sensor for glucose measurement [29].

Glucose Sensors

Another important principle that can be used in fiber-optic sensors for measurements of high sensitivity and specificity is the concept of competitive binding. This was first described by Schultz, Mansouri, and Goldstein [29] to construct a glucose sensor. In their unique sensor, the analyte (glucose) competes for binding sites on a substrate (the lectin concanavalin A) with a fluorescent indicator-tagged polymer [fluorescein isothiocyanate (FITC)-dextran]. The sensor, which is illustrated in Fig. 12.8, is arranged so that the substrate is fixed in a position out of the optical path of the fiber end. The substrate is bound to the inner wall of a glucose-permeable hollow fiber tubing (300 μ O.D.×200 μ I.D.) and fastened to the end of an optical fiber. The hollow fiber acts as the container and is impermeable to the large molecules of the fluorescent indicator. The light beam that extends from the fiber "sees" only the unbound indictor in solution inside the hollow fiber but not the indicator bound on the container wall. Excitation light passes through the fiber and into the solution, fluorescing the unbound indicator, and the fluorescent light passes back along the same fiber to a measuring system. The fluorescent indicator and the glucose are in competitive binding equilibrium with the substrate. The interior glucose concentration equilibrates with its concentration exterior to the probe. If the glucose concentration increases, the indicator is driven off the substrate to increase the concentration of the indicator. Thus, fluorescence intensity as seen by the optical fiber follows the glucose concentration.

The response time of the sensor was found to be about 5 min. *In vivo* studies demonstrated fairly close correspondence between the sensor output and actual blood glucose levels. A time lag of about 5 min was found and is believed to be due to the diffusion of glucose across the hollow fiber membrane and the diffusion of FTIC-dextran within the tubing.

In principle, the concept of competitive binding can be applied to any analysis for which a specific reaction can be devised. However, long-term stability of these sensors remains the major limiting factor that needs to be solved.

Immunosensors

Immunologic techniques offer outstanding selectivity and sensitivity through the process of antibody-antigen interaction. This is the primary recognition mechanism by which the immune system detects and fights foreign matter and has therefore allowed the measurement of many important compounds at trace levels in complex biologic samples.

In principle, it is possible to design competitive binding optical sensors utilizing immobilized antibodies as selective reagents and detecting the displacement of a labeled antigen by the analyte. Therefore, antibody-based immunologic optical systems have been the subject of considerable research in the past

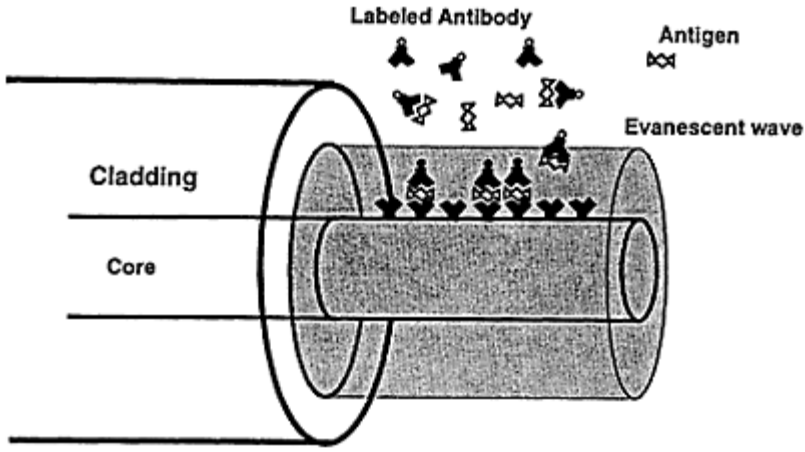


FIGURE 12.9 Basic principle of a fiber-optic antigen-antibody sensor [33].

few years [30–34]. In practice, however, the strong binding of antigens to antibodies and vice versa causes difficulties in constructing reversible sensors with fast dynamic responses.

Several immunologic sensors based on fiber-optic waveguides have been demonstrated for monitoring antibody-antigen reactions. Typically, several centimeters of cladding are removed along the fiber's distal end, and the recognition antibodies are immobilized on the exposed core surface. These antibodies bind fluorophore-antigen complexes within the evanescent wave as illustrated in Fig. 12.9. The fluorescent signal excited within the evanescent wave is then transmitted through the cladded fiber to a fluorimeter for processing.

Experimental studies have indicated that immunologic optical sensors can generally detect micromolar and even picomolar concentrations. However, the major obstacle that must be overcome to achieve high sensitivity in immunologic optical sensors is the nonspecific binding of immobilized antibodies.

References

1. Polanyi ML, Heir RM. 1962. In vivo oximeter with fast dynamic response. *Rev Sci Instrum* 33:1050.
2. Bornzin GA, Mendelson Y, Moran BL, etal. 1987. Measuring oxygen saturation and hematocrit using a fiberoptic catheter. *Proc 9th Ann Conf Eng Med Bio Soc* 807–809.
3. Mendelson Y, Galvin JJ, Wang Y. 1990. In vitro evaluation of a dual oxygen saturation/hematocrit intravascular fiberoptic catheter. *Biomed Instrum Tech* 24:199.
4. Mendelson Y. 1992. Pulse oximetry: Theory and application for noninvasive monitoring. *Clin Chem* 28(9):1601.

5. Aoyagi T, Kishi M, Yamaguchi K, et al. 1974. Improvement of the earpiece oximeter. *Jpn Soc Med Electron Biomed Eng* 90–91.
6. Yoshiya I, Shimada Y, Tanaka K. 1980. Spectrophotometric monitoring of arterial oxygen saturation in the fingertip. *Med Biol Eng Comput* 18:27.
7. Mendelson Y, Solomita MV. 1992. The feasibility of Spectrophotometric measurements of arterial oxygen saturation from the scalp utilizing noninvasive skin reflectance pulse oximetry. *Biomed Instrum Technol* 26:215.
8. Mendelson Y, McGinn MJ. 1991. Skin reflectance pulse oximetry: In vivo measurements from the forearm and calf. *J Clin Monit* 7:7.
9. Chance B, Leigh H, Miyake H, et al. 1988. Comparison of time resolved and un-resolved measurements of deoxyhemoglobin in brain. *Proc Nat Acad Sci* 85:4971.
10. Jobsis FF, Keizer JH, LaManna JC, et al. 1977. Reflection spectrophotometry of cytochrome aa3 in vivo. *Appl Physiol: Respirat Environ Excerc Physiol* 43(5):858.
11. Kurth CD, Steven IM, Benaron D, et al. 1993. Near-infrared monitoring of the cerebral circulation. *J Clin Monit* 9:163.
12. Lübbers DW, Opitz N. 1975. The pCO_2/pO_2-optode: A new probe for measurement of pCO_2 or pO_2 in fluids and gases. *Z Naturforsch C: Biosci* 30C:532.
13. Clark CL, O'Brien J, McCulloch J, etal. 1986. Early clinical experience with GasStat. *J Extra Corporeal Technol* 18:185.
14. Hill AG, Groom RC, Vinansky RP, et al. 1985. On-line or off-line blood gas analysis: Cost vs. time vs. accuracy. *Proc Am Acad Cardiovasc Perfusion* 6:148.
15. Siggaard-Andersen O, Gothgen IH, Wimberley, et al. 1988. Evaluation of the GasStat fluorescence sensors for continuous measurement of pH, pCO_2 and pO_3 during CPB and hypothermia. *Scand J Clin Lab Invest* 48(Suppl. 189):77.
16. Zimmerman JL, Dellinger RP. 1993. Initial evaluation of a new intra-arterial blood gas system in humans. *Crit Care Med* 21(4):495.
17. Gottlieb A. 1992. The optical measurement of blood gases—Approaches, problems and trends: Fiber optic medical and fluorescent sensors and applications. *Proc SPIE* 1648:4.
18. Barker SL, Hyatt J. 1991. Continuous measurement of intraarterial pHa, $PaCO_2$, and PaO_2 in the operation room. *Anesth Analg* 73:43.
19. Larson CP, Divers GA, Riccitelli SD. 1991. Continuous monitoring of PaO_2 and $PaCO_2$ in surgical patients. *Abstr Crit Care Med* 19:525.
20. Peterson JI, Goldstein SR, Fitzgerald RV. 1980. Fiber optic pH probe for physiological use. *Anal Chem* 52:864.
21. Markle DR, McGuire DA, Goldstein SR, et al: 1981. A pH measurement system for use in tissue and blood, employing miniature fiber optic probes, In DC Viano (ed), *Advances in Bioengineering*, p 123, New York, American Society of Mechanical Engineers.
22. Wolfbeis OS, Furlinger E, Kroneis H, et al. 1983. Fluorimeter analysis: 1. A study on fluorescent indicators for measuring near neutral (physiological) pH values. *Fresenius' Z Anal Chem* 314:119.
23. Wolfbeis OS, Offenbacher H. 1986. Fluorescence sensor for monitoring ionic strength and physiological pH values. *Sens Actuators* 9:85.
24. Opitz N, Lübbers DW. 1983. New fluorescence photomatrical techniques for simultaneous and continuous measurements of ionic strength and hydrogen ion activities. *Sens Actuators* 4:473.
25. Vurek GG, Feustel PJ, Severinghaus JW. 1983. A fiber optic pCO_2 sensor. *Ann Biomed Eng* 11:499.
26. Zhujun Z, Seitz WR. 1984. A carbon dioxide sensor based on fluorescence. *Anal Chim Acta* 160:305.
27. Peterson JI, Fitzgerald RV, Buckhold DK. 1984. Fiber-optic probe for in vivo measurements of oxygen partial pressure. *Anal Chem* 56:62.
28. Wolfbeis OS, Posch HE, Kroneis HW. 1985. Fiber optical fluorosensor for determination of halothane and/or oxygen. *Anal Chem* 57:2556.

29. Schultz JS, Mansouri S, Goldstein IJ. 1982. Affinity sensor: A new technique for developing implantable sensors for glucose and other metabolites. *Diabetes Care* 5:245.
30. Andrade JD, Vanwagenen RA, Gregonis DE, etal. 1985. Remote fiber optic biosensors based on evanescent-excited fluoro-immunoassay: Concept and progress. *IEEE Trans Electron Devices* ED-32:1175.
31. Sutherland RM, Daehne C, Place JF, et al. 1984. Optical detection of antibody-antigen reactions at a glass-liquid interface. *Clin Chem* 30:1533.
32. Hirschfeld TE, Block MJ. 1984. Fluorescent immunoassay employing optical fiber in a capillary tube. US Patent No. 4,447, 546.
33. Anderson GP, Golden JP, Ligler FS. 1993. An evanescent wave biosensor: Part I. Fluorescent signal acquisition from step-etched fiber optic probes. *IEEE Trans Biomed Eng* 41(6):578.
34. Golden JP, Anderson GP, Rabbany SY, et al. 1994. An evanescent wave biosensor: Part II. Fluorescent signal acquisition from tapered fiber optic probes. *IEEE Trans Biomed Eng* 41(6):585.

13
Bioanalytic Sensors

Richard P.Buck
University of North Carolina

13.1 Classification of Biochemical Reactions in the Context of Sensor Design and Development

13.1.1 Introduction and Definitions

Since sensors generate a measurable material property, they belong in some grouping of transducer devices. Sensors specifically contain a recognition process that is characteristic of a material sample at the molecular-chemical level, and a sensor incorporates a transduction process (step) to create a useful signal. Biomedical sensors include a whole range of devices that may be chemical sensors, physical sensors, or some kind of mixed sensor.

Chemical sensors use chemical processes in the recognition and transduction steps. Biosensors are also chemical sensors, but they use particular classes of biological recognition/transduction processes. A pure physical sensor generates and transduces a parameter that does not depend on the chemistry per se, but is a result of the sensor responding as an aggregate of point masses or charges. All these, when used in a biologic system (biomatrix), may be considered *bioanalytic* sensors without regard to the chemical, biochemical, or physical distinctions. They provide an "analytic signal of the biologic system" for some further use.

The chemical recognition process focuses on some molecular-level chemical entity, usually a kind of chemical structure. In classical analysis this structure may be a simple functional group: SiO^- in a glass electrode surface, a chromophore in an indicator dye, or a metallic surface structure, such as silver metal that recognizes Ag^+ in solution. In recent times, the biologic recognition processes have been better understood, and the general concept of recognition by *receptor* or *chemoreceptor* has come into fashion. Although these are often large molecules bound to cell membranes, they contain specific structures that permit a wide variety of different molecular recognition steps including recognition of large and small species and of charged and uncharged species. Thus, *chemoreceptor* appears in the sensor literature as a generic term for the principal entity doing the recognition. For a history and examples, see references [1–6].

0-8493-1813-0/03/$0.00+$1.50

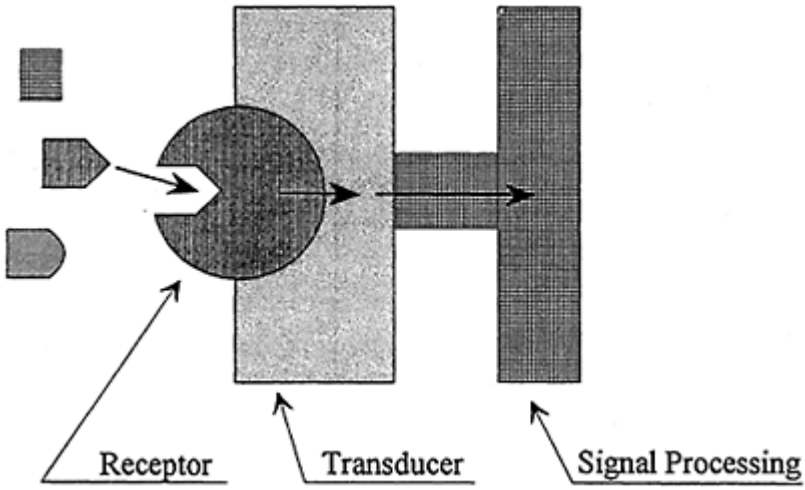


FIGURE 13.1 Generic bioanalytic sensor.

Biorecognition in biosensors has especially stressed "receptors" and their categories. Historically, application of receptors has not necessarily meant measurement directly of the receptor. Usually there are coupled chemical reactions, and the transduction has used measurement of the subsidiary products: change of pH, change of dissolved O_2,[1] generation of H_2O_2, changes of conductivity, changes of optical adsorption, and changes of temperature. Principal receptors are enzymes because of their extraordinary selectivity. Other receptors can be the more subtle species of biochemistry: antibodies, organelles, microbes, and tissue slices, not to mention the trace level "receptors" that guide ants, such as pheromones, and other unusual species. A sketch of a generic bioanalytic sensor is shown in Fig. 13.1

Classification of Recognition Reactions and Receptor Processes

The concept of *recognition* in chemistry is universal. It almost goes without saying that all chemical reactions involved recognition and selection on the basis of size, shape, and charge. For the purpose of constructing sensors, general recognition based on these factors is not usually enough. Frequently in inorganic chemistry a given ion will react indiscriminately with similar ions of the same size and charge. Changes in charge from unity to two, for example, do change the driving forces of some ionic reactions. By control of dielectric constant of phases, heterogeneous reactions can often be "tailored" to select divalent ions over monovalent ions and to select small versus large ions or vice versa.

Shape, however, has more special possibilities, and natural synthetic methods permit product control. Nature manages to use shape together with charge to build organic molecules, called enzymes, that have acquired remarkable selectivity. It is in the realm of biochemistry that these natural constructions are investigated and catalogued. Biochemistry books list large numbers of enzymes and other selective materials that

direct chemical reactions. Many of these have been tried as the basis of selective sensors for bioanalytic and biomedical purposes. The list in Table 13.1 shows how some of the materials can be grouped into lists according to function and to analytic substrate, both organic and inorganic. The principles seem general, so there is no reason to discriminate against the inorganic substrates in favor or the organic substrates. All can be used in biomedical analysis.

[1] Additional information on these types of sensors can be found in Chapter 50 and Appendix A.

TABLE 13.1 Recognition Reactions and Receptor Processes

1. Insoluble salt-based sensors		
a.	$S^+ + R^- \rightarrow I$(insoluble salt)	
Ion exchange with crystalline SR (homogeneous or heterogeneous crytals)		
	chemical signal S^{+n}	receptor R^{-n}
	inorganic cations	inorganic anions
examples:	Ag^+, Hg_2^{2+}, Pb^{2+}, Cd^{2+}, Cu^{2+}	$S^{=}$, Se^{2-}, SCN^-, I^-, Br^-, Cl^-
b.	$S^{-n} + R^{+n} \rightarrow$ SR (insoluble salt)	
Ion exchange with crystalline SR (homogeneous or heterogeneous crystals)		
	chemical signal S^{-n}	receptor R^{+n}
	inorganic anions	inorganic cations
examples:	F^-, $S^=$, Se^{2-}, SCN^-, I^-, Br^-, Cl^-	LaF_3^+, Ag^+, Hg_2^{2+}, Pb^{2+}, Cd^{2+}, Cu^{2+}
2. Solid ion exchanges		
a.	$S^{+n} + R^{-n}$ (sites) $\rightarrow S^{+n}R^{-n} = SR$ (in ion exchanger phase)	
Ion exchange with synthetic ion exchangers containing negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	H^+, Na^+, K^+	silicate glass $Si\text{-}O^-$
	H^+, Na^+, K^+, other M^{+n}	synthetic sulfonated, phosphorylated, EDTA-substituted polystyrenes
b.	$S^{-n} + R^{+n}$ (sites) $\rightarrow S^{-n}R^{+n} = SR$ (in ion exchanger phase)	
Ion exchange with synthetic ion exchangers containing positive fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{-n}	receptor R^{+n}
	organic and inorganic ions	organic and inorganic ion sites
examples:	hydrophobic anions	quaternized polystyrene
3. Liquid ion exchanger sensors with electrostatic selection		
a.	$S^{+n} + R^{-n}$ (sites) $\rightarrow S^{+n}R^{-n} = SR$ (in ion exchanger phase)	
Plasticized, passive membranes containing mobile trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	Ca^{2+}	diester of phosphoric acid or monoester of a phosphonic acid
	M^{+n}	dinonylnaphthalene sulfonate and other organic, hydrophobic anions
	$R_1R_2R_3R_4N^+$ and bis-Quaternary Cations cationic drugs tetrasubstituted arsonium$^+$	tetraphenylborate anion or substituted derivatives
b.	$S^{-n} + R^{+n}$ (sites) $\rightarrow S^{-n}R^{+n} = SR$ (in ion exchanger phase)	
Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{-n}	receptor R^{+n}
	inorganic and organic ions	inorganic and organic sites
examples:	anions, simple Cl^-, Br^-, ClO_4^-	quaternary ammonium cations: e.g. tridodecylmethyl-ammonium
	anions, complex, drugs	quaternary ammonium cations: e.g. tridodecylmethyl-ammonium
4. Liquid ion exchanger sensors with neutral (or charged) carrier selection		
a.	$S^{+n} + X$ and R^{-n} (sites) $\rightarrow S^{+n}X\ R^{-n} = SXR$ (in ion exchanger phase)	
Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	Ca^{2+}	X = synthetic ionophore complexing agent selective to Ca^{2+}

Recognition Reactions and Receptor Processes

		R^{-n} usually a substituted tetra phenylborate salt
	Na^+, K^+, H^+	X = selective ionophore complexing agent
b.	S^{-n} + X and R'^{+n} (sites)	$1S^{-n}X\ R'^{+n}$ = SXR (in ion exchanger phase)
Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)		
	chemical signal S^{-n}	receptor R'^{+n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	HPO_4^{2-}	R'^{+n} = quaternary ammonium salt
		X = synthetic ionophore complexing agent; aryl organotin compound or suggested cyclic polyamido-polyamines
	HCO_3^-	X = synthetic ionophore: trifluoro acetophenone
	Cl^-	X = aliphatic organotin compound
5. Bioaffinity sensors based on change of local electron densities		
		S + R 1SR
	chemical signal S	receptor R
	protein	dyes
	saccharide	lectin
	glycoprotein	
	substrate	enzyme
	inhibitor	
		Transferases
		Hydrolases (peptidases, esterases, etc.)
		Lyases
		Isomerases
		Ligases
	prosthetic group	apoenzyme
	antigen	antibody
	hormone	"receptor"
	substrate analogue	transport system
6. Metabolism sensors based on substrate consumption and product formation		
		$S + R\ 1SR \rightarrow P + R$
	chemical signal S	receptor R
	substrate	enzyme
examples:	lactate (SH_2)	hydrogenases catalyze hydrogen transfer from S to acceptor A (not molecular oxygen!) reversibly pyruvate + NADH + H^+ using lactate dehydrogenase
	$SH_2 + A\ 1S + AH_2$ lactate + NAD^+	
	glucose (SH_2)	
	$SH_2 + \frac{1}{2}O_2\ 1S + H_2O$ or	oxidases catalyze hydrogen transfer to molecular oxygen using glucose oxidase
	$SH_2 + O_2\ 1S + H_2O_2$	
	glucose + O_2 1gluconolactone + H_2O_2	
	reducing agents (S)	peroxidases catalyze oxidation of a substrate by H_2O_2 using horseradish peroxidase
	$2S + 2H^+ + H_2O_2\ 12S^+ + 2H_2O$	
	$Fe^{2+} + H_2O_2 + 2H^+\ 1Fe^{3+} + 2H_2O$	
	reducing agents	oxygenates catalyze substrate oxidations by molecular O_2
	L-lactate + O_2 1acetate + CO_2 + H_2O	
	cofactor	organelle
	inhibitor	microbe
	activator	tissue slice
	enzyme activity	
7. Coupled and hybrid systems using sequences, competition, anti-interference and amplification concepts and reactions.		
8. Biomimetic sensors		
	chemical signal S	receptor R
	sound	carrier-enzyme
	stress	
	light	

Source: Adapted from [2, 6].

13.2 Classification of Transduction Processes—Detection Methods

Some years ago, the engineering community addressed the topic of sensor classification—Richard M.White in IEEE Trans. Ultra., Ferro., Freq. Control (UFFC), UFFC-34 (1987) 124, and Wen H.Ko in IEEE/EMBS Symposium Abstract T.1.1 84CH2068–5 (1984). It is interesting because the physical and chemical properties are given equal weight. There are many ideas given here that remain without embodiment. This list is reproduced as Table 13.2. Of particular interest in this section are "detection means used in sensors" and "sensor conversion phenomena." At present the principal transduction schemes use electrochemical, optical, and thermal detection effects and principles.

TABLE 13.2 Detection Means and Conversion Phenomena Used in Sensors

Detection means
Biologic
Chemical
Electric, magnetic, or electromagnetic wave
Heat, temperature
Mechanical displacement of wave
Radioactivity, radiation
Other
Conversion phenomena
Biologic
Biochemical transformation
Physical transformation
Effect on test organism
Spectroscopy
Other
Chemical
Chemical transformation
Physical transformation
Electrochemical process
Spectroscopy
Other
Physical

Thermoelectric
Photoelectric
Photomagnetic
Magnetoelectric
Elastomagnetic
Thermoelastic
Elastoelectric
Thermomagnetic
Thermooptic
Photoelastic
Other

Calorimetric, Thermometric, and Pyroelectric Transducers

Especially useful for enzymatic reactions, the generation of heat (enthalpy change) can be used easily and generally. The enzyme provides the selectivity and the reaction enthalpy cannot be confused with other reactions from species in a typical biologic mixture. The ideal aim is to measure total evolved heat, i.e., to perform a calorimetric measurement. In real systems there is always heat loss, i.e., heat is conducted away by the sample and sample container so that the process cannot be adiabatic as required for a total heat evolution measurement. As a result, temperature difference before and after evolution is measured most often. It has to be assumed that the heat capacity of the specimen and container is constant over the small temperature range usually measured.

The simplest transducer is a thermometer coated with the enzyme that permits the selected reaction to proceed. Thermistors are used rather than thermometers or thermocouples. The change of resistance of certain oxides is much greater than the change of length of a mercury column or the microvolt changes of thermocouple junctions.

Pyroelectric heat flow transducers are relatively new. Heat flows from a heated region to a lower temperature region, controlled to occur in one dimension. The lower-temperature side can be coated with an enzyme. When the substrate is converted, the lower-temperature side is warmed. The Pyroelectric material is from a category of materials that develops a spontaneous voltage difference in a thermal gradient. If the gradient is disturbed by evolution or adsorption of heat, the voltage temporarily changes.

In biomedical sensing, some of the solid-state devices based on thermal sensing cannot be used effectively. The reason is that the sensor itself has to be heated or is heated quite hot by catalytic surface reactions. Thus pellistors (oxides with catalytic surfaces and embedded platinum wire thermometer), chemiresistors, and “Figaro” sensor “smoke” detectors have not found many biologic applications.

Optical, Optoelectronic Transducers

Most optical detection systems for sensors are small, i.e., they occupy a small region of space because the sample size and volume are themselves small. This means that common absorption spectrophotometers and photofluorometers are not used with their conventional sample-containing cells, or with their conventional beam-handling systems. Instead light-conducting *optical fibers* are used to connect the sample with the more remote monochromator and optical readout system. The techniques still remain absorption spectrophotometry, fluorimetry including fluorescence quenching, and reflectometry.

The most widely published optical sensors use a miniature reagent contained or immobilized at the tip of an optical fiber. In most systems, a permselective membrane coating allows the detected species to penetrate the dye region. The corresponding absorption change, usually at a sensitive externally preset wavelength, is changed and correlated with the sample concentration. Similarly, fluorescence can be stimulated by the higher-frequency external light source and the lower-frequency emission detected. Some configurations are illustrated in references [1,2]. Fluorimetric detection of coenzyme A, NAD+/ NADH, is involved in many so-called pyridine-linked enzyme systems. The fluorescence of NADH contained or immobilized can be a convenient way to follow these reactions. Optodes, miniature encapsulated dyes, can be placed *in vivo*. Their fluorescence can be enhanced or quenched and used to detect acidity, oxygen, and other species.

A subtle form of optical transduction uses the “peeled” optical fiber as a multiple reflectance cell. The normal fiber core glass has a refractive index greater than that of the exterior coating; there is a range of angles of entry to the fiber so that *all* the light beam remains inside the core. If the coating is removed and materials of lower index of refraction are coated on the exterior surface, there can be absorption by multiple reflections, since the evanescent wave can penetrate the coating. Chemical reagent can be added externally to create selective layers on the optical fiber.

Ellipsometry is a reflectance technique that depends on the optical constants and thickness of surface layer. For colorless layers, a polarized light beam will change its plane of polarization upon reflection by the surface film. The thickness can sometimes be determined when optical constants are known or approximated by constants of the bulk material. Antibody-antigen surface reaction can be detected this way.

Piezoelectric Transducers

Cut quartz crystals have characteristic modes of vibration that can be induced by painting electrodes on the opposite surfaces and applying a megaHertz ac voltage. The frequency is searched until the crystal goes into a resonance. The resonant frequency is very stable. It is a property of the material and maintains a value to a few parts per hundred million. When the surface is coated with a stiff mass, the frequency is altered. The shift in frequency is directly related to the surface mass for thin, stiff layers. The reaction of a

substrate with this layer changes the constants of the film and further shifts the resonant frequency. These devices can be used in air, in vacuum, or in electrolyte solutions.

Electrochemical Transducers

Electrochemical transducers are commonly used in the sensor field. The main forms of electrochemistry used are potentiometry [zero-current cell voltage (potential difference measurements)], amperometry (current measurement at constant applied voltage at the working electrode), and ac conductivity of a cell.

Potentiometric Transduction

The classical generation of an activity-sensitive voltage is spontaneous in a solution containing both nonredox ions and redox ions. Classical electrodes of types 1, 2, and 3 respond by ion exchange directly or indirectly to ions of the same material as the electrode. Inert metal electrodes (sometimes called *type 0*)—Pt, Ir, Rh, and occasionally carbon C—respond by electrons exchange from redox pairs in solution. Potential differences are interfacial and reflect ratios of activities of oxidized to reduced forms.

Amperometric Transduction

For dissolved species that can exchange electrons with an inert electrode, it is possible to force the transfer in one direction by applying a voltage very oxidizing (anodic) or reducing (cathodic). When the voltage is fixed, the species will be, by definition, out of equilibrium with the electrode at its present applied voltage. Locally, the species (regardless of charge) will oxidize or reduce by moving from bulk solution to the electrode surface where they react. Ions do not move like electrons. Rather they diffuse from high to low concentration and do not usually move by drift or migration. The reason is that the electrolytes in solutions are at high concentrations, and the electric field is virtually eliminated from the bulk. The field drops through the first 1000 Angstroms at the electrode surface. The concentration of the moving species is from high concentration in bulk to zero at the electrode surface where it reacts. This process is called *concentration polarization*. The current flowing is limited by mass transport and so is proportional to the bulk concentration.

Conductometric Transducers

Ac conductivity (impedance) can be purely resistive when the frequency is picked to be about 1000 to 10,000 Hz. In this range the transport of ions is sufficiently slow that they never lose their uniform concentration. They simply quiver in space and carry current forward and backward each half cycle. In the lower and higher frequencies, the cell capacitance can become involved, but this effect is to be avoided.

13.3 Tables of Sensors from the Literature

The longest and most consistently complete references to the chemical sensor field is the review issue of *Analytical Chemistry Journal*. In the 1970s and 1980s these appeared in the April issue, but more recently they appear in the June issue. The editors are Jiri Janata and various colleagues [7–10]. Note all possible

TABLE 13.3 Chemical Sensors and Properties Documented in the Literature

1.	General topics including items II-V; selectivity, fabrication, data processing
II.	Thermal sensors
III.	Mass sensors
	Gas sensors
	Liquid sensors
IV.	Electrochemical sensors
	Potentiometric sensors
	Reference electrodes
	Biomedical electrodes
	Applications to cations, anions
	Coated wire/hybrids
	ISFETs and related
	Biosensors
	Gas sensors
	Amperometric sensors
	Modified electrodes
	Gas sensors
	Biosensors
	Direct electron transfer
	Mediated electron transfer
	Biomedical
	Conductimetric sensors
	Semiconducting oxide sensors
	Zinc oxide-based

	Chemiresistors
	Dielectrometers
V.	Optical sensors
	Liquid sensors
	Biosensors
	Gas sensors

TABLE 13.4 Books and Long Reviews Keyed to Items in Table 13 (Reviewed since 1988 in reverse time sequence)

1. Yamauchi S (ed). 1992. *Chemical Sensor Technology*, Vol 4, Tokyo, Kodansha Ltd.

   Flores JR, Lorenzo E. 1992. Amperometric biosensors, In MR Smyth , JG Vos (eds), *Comprehensive Analytical Chemistry*, Amsterdam, Elsevier

   Vaihinger S, Goepel W. 1991. Multicomponent analysis in chemical sensing. In W Goepel, J Hesse, J Zemel (eds), *Sensors*, Vol 2, Part 1, pp 191–237, Weinheim, Germany, VCH Publishers

   Wise DL (ed). 1991. *Bioinstrumentation and Biosensors*, New York, Marcel Dekker

   Scheller F, Schubert F. 1989. *Biosensors*, Basel, Switzerland, Birkhauser Verlag, see also [2].

   Madou M, Morrison SR. 1989. *Chemical Sensing with Solid State Devices*, New York, Academic Press.

   Janata J. 1989. *Principles of Chemical Sensors*, New York, Plenum Press.

   Edmonds TE (ed). 1988. *Chemical Sensors*, Glasgow, Blackie.

   Yoda K. 1988. Immobilized enzyme cells. *Methods Enzymology*, 137:61.

   Turner APF, Karube I, Wilson GS (eds). 1987. *Biosensors: Fundamentals and Applications*, Oxford, Oxford University Press.

   Seiyama T (ed). 1986. *Chemical Sensor Technology*, Tokyo, Kodansha Ltd.

II. Thermal Sensor

   There are extensive research and application papers and these are mentioned in books listed under I. However, the up-to-date lists of papers are given in references 7–10.

III. Mass Sensors

   There are extensive research and application papers and these are mentioned in books listed under I. However, the up-to-date lists of papers are given in references 7–10. Fundamentals of this rapidly expanding field are recently reviewed:

   Buttry DA, Ward MD. 1992. Measurement of interfacial processes at electrode surfaces with the electrochemical quartz crystal microbalance, *Chemical Reviews* 92:1355.

Grate JW, Martin SJ, White RM. 1993. Acoustic wave microsensors, Part 1, *Analyt Chem* 65:940A; part 2, *Analyt Chem* 65:987A.

Ricco AT. 1994. SAW Chemical sensors, *The Electrochemical Society Interface Winter:* 38–44.

IVA. Electrochemical Sensors—Liquid Samples

Scheller F, Schmid RD (eds). 1992. Biosensors: *Fundamentals, Technologies and Applications,* GBF Monograph Series, New York, VCH Publishers.

Erbach R, Vogel A, Hoffmann B. 1992. Ion-sensitive field-effect structures with Langmuir-Blodgett membranes. In F Scheller, RD Schmid (eds). *Biosensors: Fundamentals, Technologies, and Applications,* GBF Monograph 17, pp 353–357, New York, VCH Publishers.

Ho May YK, Rechnitz GA. 1992. An introduction to biosensors, In RM Nakamura, Y Kasahara, GA Rechnitz (eds), *Immunochemical Assays and Biosensors Technology,* pp 275–291, Washington, DC, American Society Microbiology.

Mattiasson B, Haakanson H. Immunochemically-based assays for process control, 1992. *Advances in Biochemical Engineering and Biotechnology* 46:81.

Maas AH, Sprokholt R. 1990. Proposed IFCC Recommendations for electrolyte measurements with ISEs in clinical chemistry, In A Ivaska, A Lewenstam, R Sara (eds), *Contemporary Electroanalytical Chemistry, Proceedings of the ElectroFinnAnalysis International Conference on Electroanalytical Chemistry,* pp 311–315, New York, Plenum.

Vanrolleghem P, Dries D, Verstreate W. RODTOX: Biosensor for rapid determination of the biochemical oxygen demand, 1990. In C Christiansen, L Munck, J Villadsen (eds), *Proceedings of the 5th European Congress Biotechnology,* Vol 1, pp 161–164, Copenhagen, Denmark, Munksgaard.

Cronenberg C, Van den Heuvel H, Van den Hauw M, Van Groen B. Development of glucose microelectrodes for measurements in biofilms, 1990. In C Christiansen, L Munck, J Villadsen J (eds), *Proceedings of the 5th European Congress Biotechnology,* Vol 1, pp 548–551, Copenhagen, Denmark, Munksgaard.

Wise DL (ed). 1989. *Bioinstrumentation Research, Development and Applications,* Boston, MA, Butterworth-Heinemann.

Pungor E (ed). 1989. *Ion-Selective Electrodes—Proceedings of the 5th Symposium* (Matrafured, Hungary 1988), Oxford, Pergamon.

Wang J (ed). 1988. *Electrochemical Techniques in Clinical Chemistry and Laboratory Medicine,* New York, VCH Publishers.

Evans A. 1987. *Potentiometry and Ion-selective Electrodes,* New York, Wiley. Ngo TT(ed). 1987. *Electrochemical Sensors in Immunological Analysis,* New York, Plenum.

IVB. Electrochemical Sensors—Gas Samples

Sberveglieri G (ed). 1992. *Gas Sensors,* Dordrecht, The Netherlands, Kluwer.

Moseley PT, Norris JOW, Williams DE. 1991. *Technology and Mechanisms of Gas Sensors,* Bristol, U.K., Hilger.

Moseley PT, Tofield BD (eds). 1989. *Solid State Gas Sensors*, Philadelphia, Taylor and Francis, Publishers.

Books and Long Reviews Keyed to Items in Table 13 (Reviewed since 1988 in reverse time sequence)

V. Optical Sensors

Coulet PR, Blum LJ. Luminescence in biosensor design, 1991. In DL Wise, LB Wingard, Jr (eds). *Biosensors with Fiberoptics*, pp 293–324, Clifton, N.J., Humana.

Wolfbeis OS. 1991. Spectroscopic techniques, In OS Wolfbeis (ed). *Fiber Optic Chemical Sensors and Biosensors*, Vol 1, pp 25–60. Boca Raton, Fla., CRC Press.

Wolfbeis OS. 1987. Fibre-optic sensors for chemical parameters of interest in biotechnology, In RD Schmidt (ed). GBF (Gesellschaft fur Biotechnologische Forschung) *Monogr. Series*, Vol 10, pp 197–206, New York, VCH Publishers.

or imaginable sensors have been made according to the list in Table 13.2. A more realistic table can be constructed from the existing literature that describes actual devices. This list is Table 13.3. Book references are listed in Table 13.4 in reverse time order to about 1986. This list covers most of the major source books and many of the symposium proceedings volumes. The reviews [7–10] are a principal source of references to the published research literature.

13.4 Applications of Microelectronics in Sensor Fabrication

The reviews of sensors since 1988 cover fabrication papers and microfabrication methods and examples [7–10]. A recent review by two of the few *chemical* sensor scientists (chemical engineers) who also operate a microfabrication laboratory is C.C.Liu, Z.-R.Zhang. 1992. Research and development of chemical sensors using microfabrication techniques. *Selective Electrode* 14:147.

References

1. Janata J. 1989. *Principles of Chemical Sensors*, New York, Plenum.
2. Scheller F, Schubert F. 1989. *Biosensors*, #18 in Advances in Research Technologies (Beitrage zur Forschungstechnologie), Berlin, Akademie-Verlag, Amsterdam, Elsevier (English translation)
3. Turner APF, Karube I, Wilson GS. 1987. *Biosensors: Fundamentals and Applications*, Oxford, Oxford University Press.
4. Hall EAH. 1990. *Biosensors*, Milton Keynes, England, Open University Press.
5. Eddoes MJ. 1990. Theoretical methods for analyzing biosensor performance. In AEG Cass (ed), *Biosensor—A Practical Approach*, Oxford, IRL Press at Oxford University, Ch. 9, pp 211–262.
6. Cosofret VV, Buck RP. 1992. *Pharmaceutical Applications of Membrane Sensors*, Boca Raton, Fla, CRC Press.
7. Janata J, Bezegh A. 1988. Chemical sensors, *Analyt Chem* 60:62R.

8. Janata J. 1990. Chemical sensors, *Analyt Chem* 62:33R.
9. Janata J. 1992. Chemical sensors, *Analyt Chem* 66:196R.
10. Janata J, Josowicz M, DeVaney M. 1994. Chemical sensors, *Analyt Chem* 66:207R.

Appendix 1
Historical Perspective: The Electrocardiograph

Leslie A.Geddes
Purdue University

Recording and display of bioelectric events occupied a long time and required the development and adaptation of a variety of primitive instruments, not all of which were electronic. The first bioelectric recorder was the rheoscopic frog, consisting of a sciatic nerve and its innervated gastrocnemius muscle. So sensitive was the nerve that it could be stimulated by the beating heart or a contracting muscle; both events contracted the gastrocnemius muscle. However, such a response provided no information on the time course of these bioelectric events. Sensitive and rapidly responding indicators were essential for this purpose.

As Etienne Jules Marey [1885], champion of the graphic method, stated:

> In effect, in the field of rigorous experimentation all the sciences give a hand. Whatever is the object of these studies, that which measures a force or movement, an electrical state or a temperature, whether he be a physician, chemist, or physiologist, he has recourse to the same method and employs the same instruments.

Development of the galvanometer and the electric telegraph provided design concepts and instruments that could be adapted to the measurement of bioelectric events. For example, Thomson's reflecting telegraphic galvanometer was used by Caton [1875] to display the first electroencephalogram. Ader's telegraphic string galvanometer [1897] was modified by Einthoven [1903] to create the instrument that introduced clinical electrocardiography. Gasser and Erlanger [1922] adapted the Braun cathode-ray tube to enable recording of short-duration nerve action potentials. Garceau [1935] used the Western Union telegraphic recorder, called the Undulator to create the first direct-inking electroencephalograph. However, in the early days of bioelectricity, ingenious electrophysiologists appropriated many devices from physics and engineering to establish the existence of bioelectric phenomena.

The First Electrocardiogram

The electric activity accompanying the heartbeat was discovered with the rheoscopic frog by Kolliker and Mueller [1856]. When these investigations laid the nerve over the beating ventricle of a frog heart, the muscle twitched once and sometimes twice. Stimulation of the nerve obviously occurred with depolarization and repolarization of the ventricles. Because at that time there were no rapidly responding galvanometers, Donders [1872] recorded the twitches of the rheoscope to provide a graphic demonstration of the existence of an electrocardiographic signal.

0-8493-1813-0/03/$0.00+$1.50

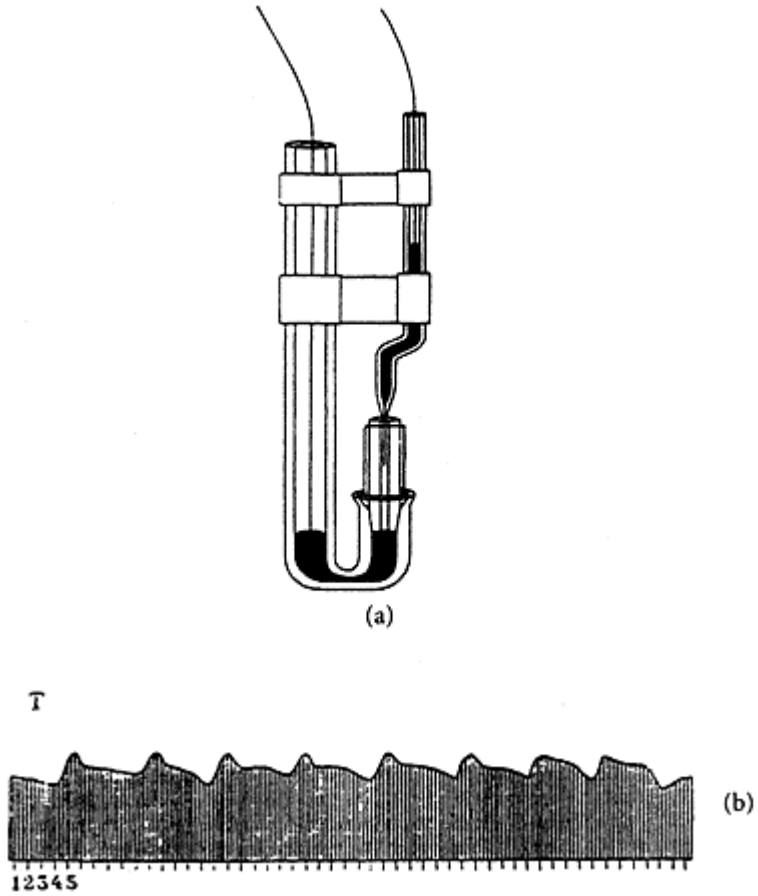


FIGURE A1.1 The capillary electrometer *(a)* used by Marey and Lippmann in 1876 and a tortoise ventricular electrogram *(b)* made with it. This is the first cardiac electrogram from a spontaneously beating heart.

Capillary Electrometer Record

The capillary electrometer was created especially for recording the electrocardiogram. The principle underlying its operation was being investigated by Lippmann, a colleague of Marey in France. The phenomenon of electrocapillarity is the change in contour of a drop of mercury in dilute sulfuric acid when a current is passed through the mercury—sulfuric acid interface. This phenomenon was put to practical use by Marey [1876], who

placed the interface in a capillary tube, transilluminated it, and recorded the contour change on a moving (falling) photographic plate. The two wires from the electrometer were connected to electrodes placed against the exposed tortoise ventricle. Figure A1.1*a* illustrates the capillary electrometer, and Fig. A1.1*b* is a reproduction of the tortoise ventricular electrogram showing what we now call R and T waves.

Rheotome Record

Probably unaware that Marey had recorded the cardiac electrogram with the capillary electrometer, Burdon-Sanderson [1879] in England used a slow-speed, d'Arsonval-type galvanometer, the rheotome [see Hoff and Geddes, 1957], and induction-coil stimulator [see Geddes et al., 1989] to reconstruct the ventricular electrogram of the frog heart; Fig. A1.2 illustrates his reconstruction, showing the R and T waves; note their similarity with those obtained by Marey in Fig. A1.1*b*.

Mammalian Electrocardiograms

When news of the capillary electrometer reached the United Kingdom, many investigators fabricated their own instruments. One of these was Waller, who used it to record the electrocardiogram of a

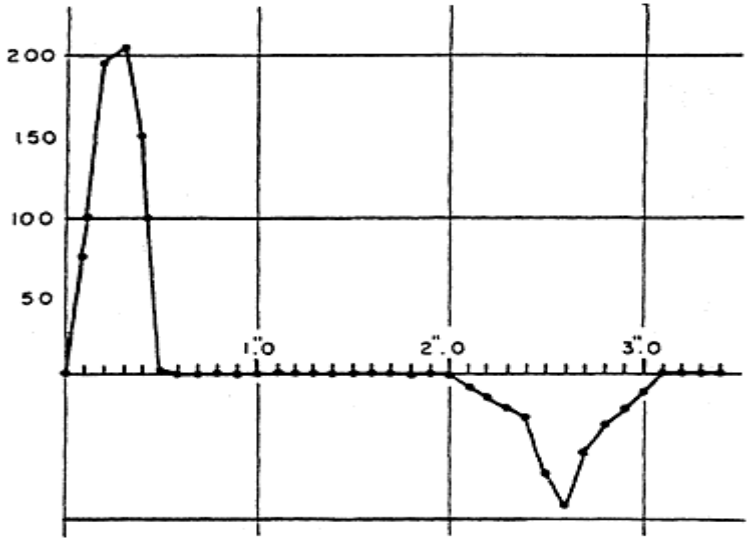


FIGURE A1.2 Burdon-Sanderson's plot of the frog cardiac electrogram. Thirty-five points were determined to

make the reconstruction. [Burdon-Sanderson and Page, 1879.]

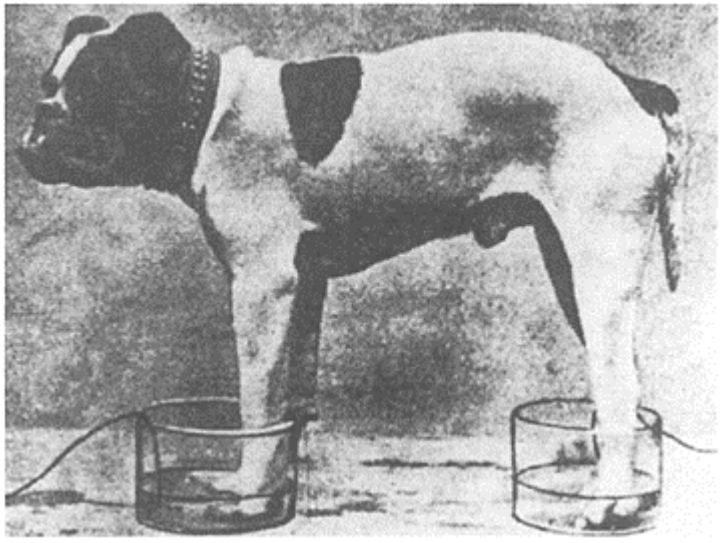


FIGURE A1.3 Waller's patient Jimmy having his EGG recorded with the capillary electrometer. (From Waller AD, Hitchcock Lectures, University of London, 1910.)

patient whom he called Jimmy. In 1910, Waller revealed the identity of Jimmy, his pet bulldog, shown in Fig. A1.3 having his EGG recorded with a forepaw and hindpaw in glass vessels containing saline and metal electrodes.

Waller [1887] obtained the first ECGs from human subjects; Fig. A1.4*a* is one of his records, which displays the apex cardiogram and the capillary electrometer record, showing R and T waves.

At that time there were no standard sites for electrode placement. Using the extremities, Waller experimented with different sites, discovering that there were favorable and unfavorable sites, i.e., sites where the amplitude was large or small; Table A1.1 summarizes his findings. From recordings made with these electrodes, Waller [1887] proposed that the heart could be represented as a dipole, as shown in Fig. A1.4*b*.

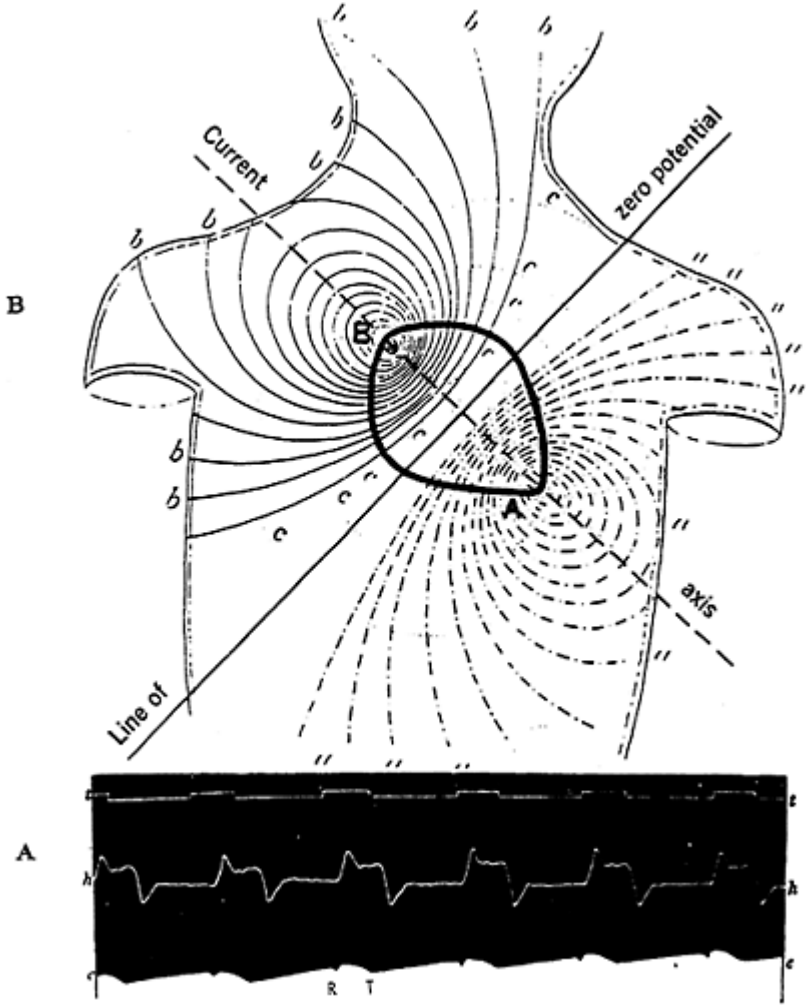


FIGURE A1.4 First human apex cardiogram and capillary electrometer EGG *(a)* and the dipole map for the heart (b), (*a* from Waller [1887]; *b* from Waller [1889].)

TABLE A1.1 Waller's Electrode Locations

The unfavorable combinations were:
Left hand and left foot
Left hand and right foot

Right foot and left foot
Mouth and right hand
The favorable combinations were:
Front of chest and back of chest
Left hand and right hand
Right hand and right foot
Right hand and left foot
Mouth and left hand
Mouth and right foot
Mouth and left foot

Corrected Capillary Electrometer Records

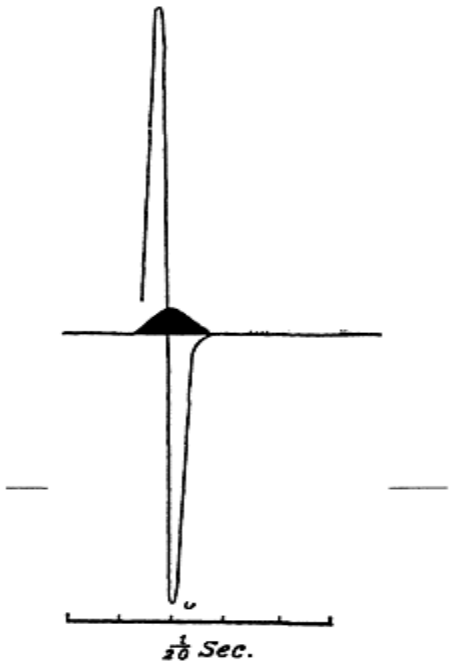


FIGURE A1.5 Capillary electrometer record *(dark hump)* and the corrected voltage-time record *(biphasic wave).*

By the 1890s, it was known that the response of the capillary electrometer was slow, and methods were developed to correct recordings to obtain a true voltage-time record. Burch [1892] in the United Kingdom developed a geometric method that used the tangent at each point along the recording. Figure A1.5 is an illustration showing a capillary electrometer record (dark hump) and the true voltage-time record (biphasic waveform).

One who was very dissatisfied with capillary-electrometer records was Einthoven in the Netherlands. He obtained a capillary electrometer record from a subject (Fig. A1.6*a*) and applied the correct method to create the EGG shown in Fig. A1.6*b*, revealing the intimate details of what he called Q and S waves, not visible in the capillary electrometer record, in which he used A, B, and C to designate what he later called P, R, and T waves.

Clinical Electrocardiography

AP1.0.1 The String Galvanometer

Einthoven [1903] set himself the task of creating a high-fidelity recorder by improving Ader's string telegraphic galvanometer. Einthoven used a silvered quarter filament as the conductor and increased the field strength surrounding the filament by using a strong electromagnet. He added a lens to focus the image of the filament onto a moving photographic surface. Thus the thick baseline for the string galvanometer recording was the image of the "string" (quartz filament). Electrocardiac current caused the silvered filament to be deflected, and its excursions, when magnified optically and recorded photographically, constituted the electrocardiogram. Figure A1.7*a* illustrates Einthoven's string galvanometer, and Fig. A1.7*b* shows a patient in hospital clothing having his EGG recorded. Measurement of calibration records published by Einthoven reveals a response time (10% to 90%) of 20 ms. The corresponding sinusoidal frequency response is 0 to 25 Hz (30% attenuation).

Figure A1.8 illustrates one of Einthoven's electromagnets (lower left) and one of his early cameras (center) and string galvanometers (right). Above the bench supporting these are mounted a bow and arrow, which were used by Einthoven to provide the tension on the quartz rod while it was being heated to the melting point, after which the arrow left the bow and extruded the rod to a long slender filament, which was later silvered to make it conducting.

Einthoven borrowed handsomely from previous work; he used Marey's recording chart speed (25 mm/ s) and Waller's bucket electrodes, as well as some of his leads. With the string galvanometer, Einthoven ushered in clinical electrocardiography in the early 1900s; soon the string galvanometer was used worldwide. The first string galvanometer appeared in the United States in 1912; it was used at the Rockefeller Institute Hospital until 1959. This instrument, which is now in the Smithsonian Institute, was designed by Horatio B.Williams, professor of physiology at the College of Physicians and Surgeons at Columbia University. Williams had spent some time with Einthoven in Leiden in 1910 and 1911. On his return to New York, Williams had Charles F. Hindle, a machinist at Columbia, construct the first American string galvanometer. Soon thereafter, the Cambridge Instrument Company took over manufacture of the Hindle instrument and made them available for sale in the United States.

Although it is clear that the concept of an electrical axis, i.e., a cardiac vector, was demonstrated by Waller's studies, it remained for Einthoven [1913] to make practical use of the concept. Einthoven

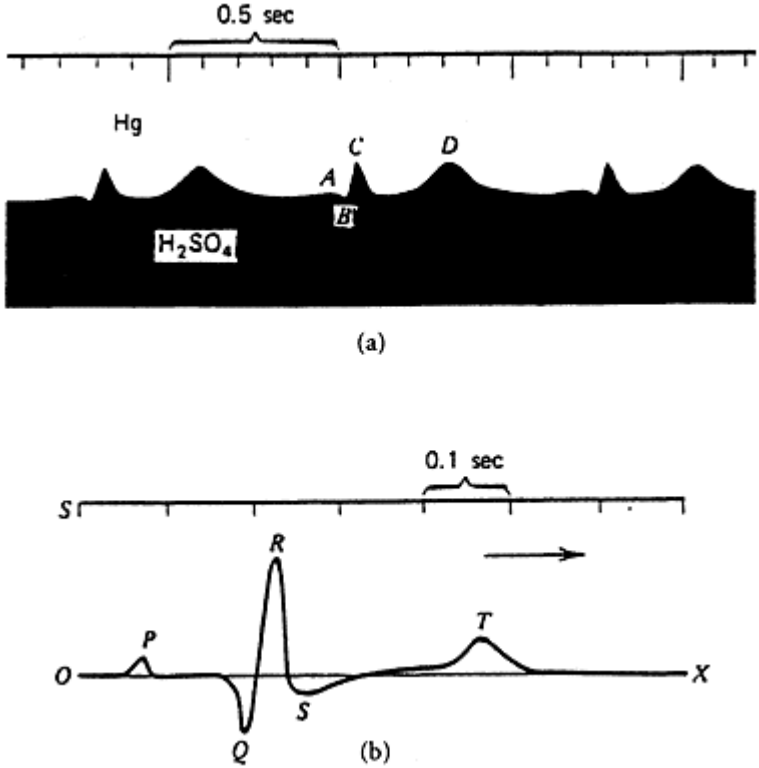


FIGURE A1.6 A capillary electrometer record *(a)* and its corrected version *(b)* presented by Einthoven to show the inadequacy of the capillary electrometer in displaying rapidly changing waveforms. (From FA Willius, 1941. TE Keys, Cardiac Classics, St. Louis, Mosby.)

postulated that the heart was at the center of an equilateral triangle, the apices of which were the right and left shoulders and the point where both legs joined the trunk. In his early studies, Einthoven used the right and left arms and both feet in saline-filled buckets as the three electrodes. Soon he found that the electrocardiogram was negligibly altered if the right foot was removed from the bucket electrode. Thus he adopted three standard leads: right and left arms and left leg (foot). He postulated that if the amplitudes of the

electrocardiographic waves are plotted on this triaxial reference frame, it is possible to calculate the magnitude and direction of an electric vector that produces these same voltages in leads I, II, and III, corresponding to the limb electrodes. He further stated that the arithmetic sum of the amplitudes in lead I plus III equals the amplitude in lead II. This is *Einthoven's law,* and the relationship is true only for an equilateral triangle reference frame [Valentinuzzi et al., 1970].

Vacuum-Tube Electrocardiograph

Not long after Einthoven described his string galvanometer, efforts were begun in the United States to create an electrocardiograph that used vacuum tubes. At that time, there were rapidly responding mirror galvanometers, as well as a limited number of vacuum tubes, despite the fact that they has been patented only a few years earlier [1907] by DeForest. According to Marvin [1954], the first discussions relative to such an instrument were held in 1917 between Steinmetz, Neuman, and Robinson of the General Electric Engineering Laboratory. The task of establishing feasibility fell to W.R.G.Baker, who assembled a unit and demonstrated its operation to those just identified. However, because of urgent wartime priorities, the project was shelved.

In 1921, General Electric reopened the issue of a vacuum-tube EGG. A second prototype was built and demonstrated to the Schenectady County Medical Association some time in 1924 by Robinson and Marvin. The instrument was used by Drs. Newman, Pardee, Mann, and Oppenheim, all physicians in New York City. Subsequently, six commercial models were made. One instrument was sent to each of the four physicians just identified; the fifth was sent to the General Electric Company Hospital; and the sixth was sent to the AM A Convention in Atlantic City in 1925. This latter instrument became a prototype for future models provided by the General Electric X-Ray Division. A U.S. patent application was filed on January 15, 1925, and the instrument was described by Mann [1930].

On December 22, 1931, a patent on the General Electric vacuum-tube ECG was granted to Marvin and Leibing; Fig. A1.9 shows the circuit diagram of the instrument, including a specimen record (*lower left,* Fig. A1.9). The instrument used three triode vacuum tubes in a single-sided, resistance-capacitance-coupled amplifier. It was battery operated, and a unique feature was a viewing screen that allowed the operator to see the motion of the galvanometer beam as it was being recorded by the camera.

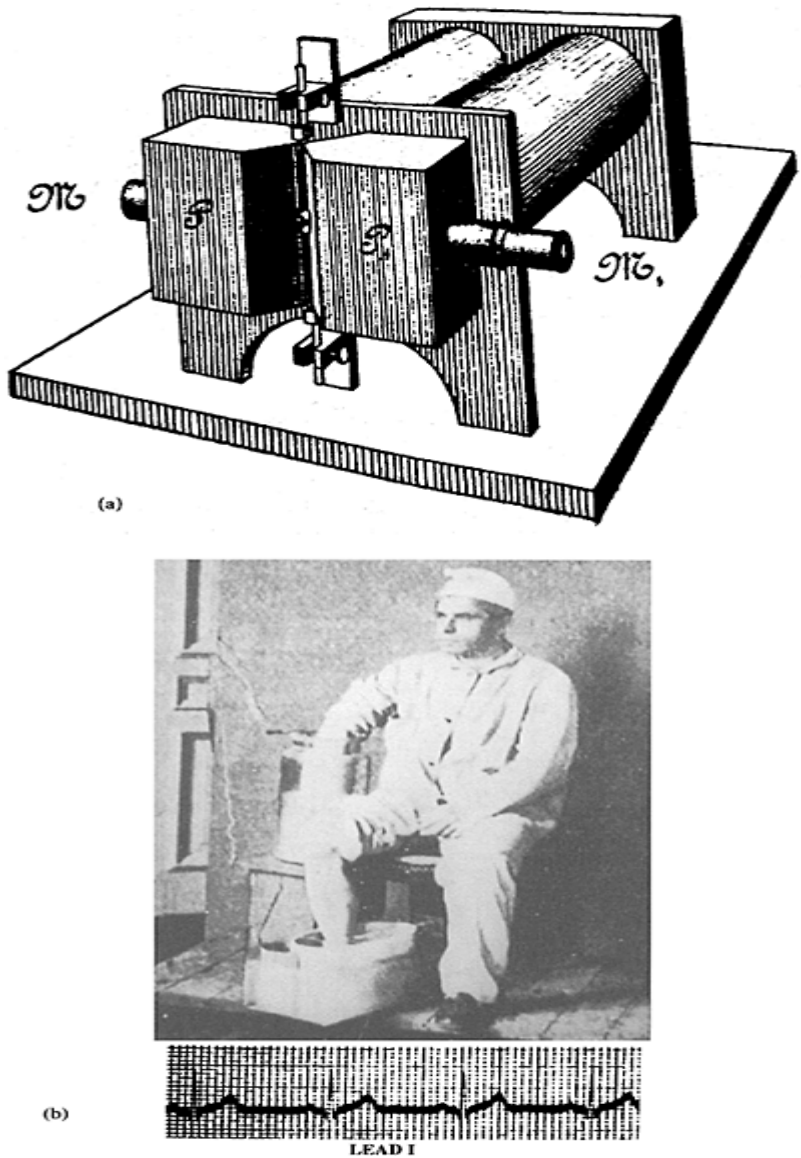


FIGURE A1.7 Einthoven's string galvanometer *(a)* and a patient having his EGG (lead 1) recorded *(b)*.

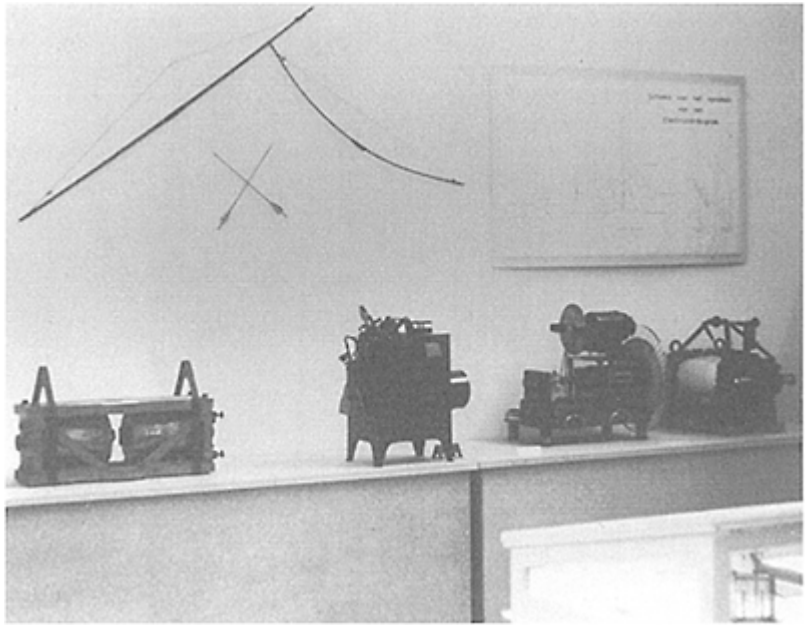


FIGURE A1.8 Some of Einthoven's early equipment. On the bench *(left)* is an electromagnet. In the center is a camera, and on the right is a string galvanometer. On the wall are a bow and arrow; the latter was used to apply force to a quartz rod which was heated, and when pulled by the arrow, created the quartz filament that was later silvered.

Between introduction of the string galvanometer and the hot-stylus recorder for ECG, attempts were made to create direct-inking ECG recorders. In a review of scientific instruments, Brian Matthews [1935] reported:

> There are two ink-writing electrocardiographs available, the author's and that of Drs. Duschel and Luthi. Both utilize a moving-iron driving unit with oil damping, a tubular pen writing on moving paper. The former has a battery-coupled amplifier. The latter gives, in effect, D.C. amplification; the potentials to be recorded are interrupted about 500 times per second by a special type of buzzer, after amplification by resistance capacity coupled valves the interrupted output is reflected by the output valve; the amplifier achieves in effect what can be done with a battery-coupled amplifier and obviates the coupling batteries. At present the speed of these direct-

recording instruments is barely adequate to show the finer details of the electrocardiogram, but they enable its main features to be recorded instantly.

Despite the instant availability of inked recordings of the ECG, those produced by the string galvanometer were superior, and it took some time for a competitor to appear. Such an instrument did appear in the form of the hot-stylus recorder.

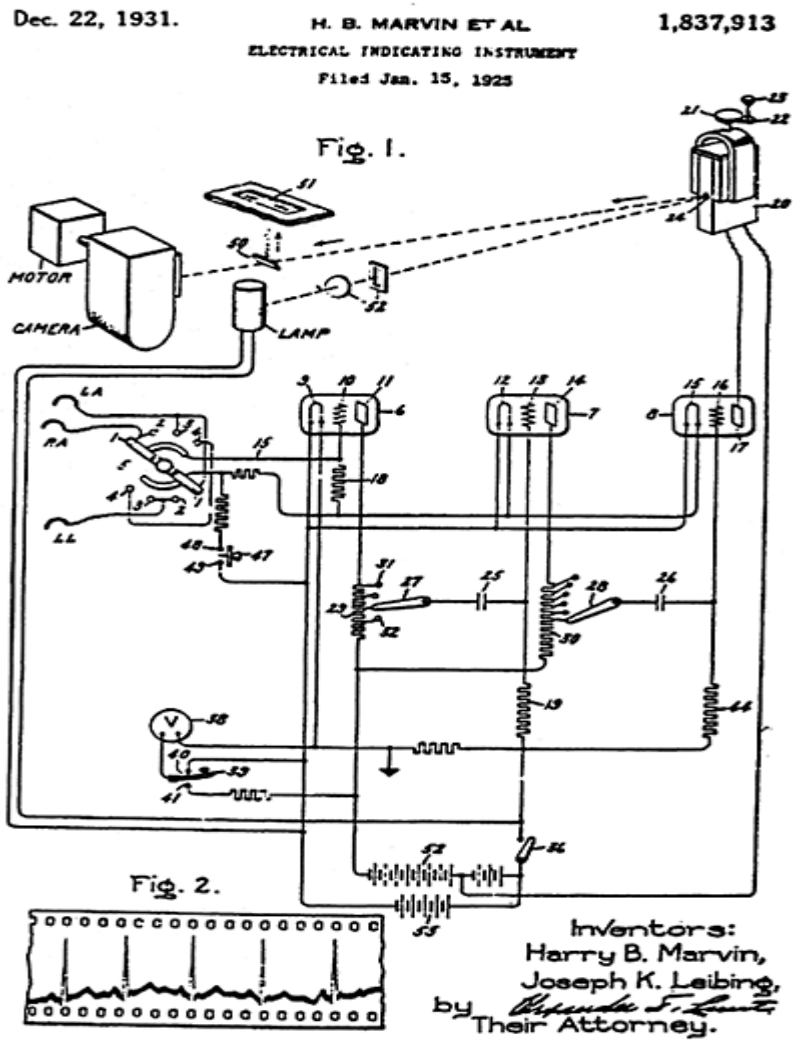


FIGURE A1.9 Circuit diagram of the first vacuum-tube EGG, patented on December 12, 1931.

Hot-Stylus Recorder

The final step toward modern electrocardiography was the introduction of the hot-stylus recorder by Haynes [1936] of the Bell Telephone Laboratories (New York). Prior to that time, there was colored, wax-coated recording paper, the wax being scraped off by a recording stylus, exposing the colored paper. Referring to the scraping method, Haynes wrote

> However, the method is not adaptable to many types of recording instruments because of the large amount of friction between the recording stylus and paper arising from the pressure necessary to engrave the wax.
>
> This pressure can be removed and the friction largely eliminated by the use of a special stylus consisting essentially of a small electric heating coil situated close to the end of a pointed rod in such a way that the temperature of the point may be raised to about 80°C. The point is then capable of melting wax and engraving its surface with only a very small fraction of the pressure before necessary.
>
> The described stylus, when used with waxed recording paper, will provide a means of obtaining a permanent record without the use of pen and ink which is adaptable to the most sensitive recording instruments.

Following the end of World War II, vacuum-tube electrocardiographs with heated-stylus recorders became very popular; they are still in use today. However, the heritage of a thick baseline, derived from the string-galvanometer days, had to be preserved for some time because clinicians objected to a thin baseline. It took many years for the hot-stylus baseline to be narrowed without protest.

References

Ader M. 1987. Sur un nouvel appareil enregistreur pour cables sousmarins. *C R Acad Sci* 124:1440.

Burch GJ. 1892. On the time relations of the excursions of the capillary electrometer. *Philos Trans R Soc* (Lond) 83A:81.

Burch GJ. 1892. On a method of determining the value of rapid variations of potential by means of the capillary electrometer communicated by J.B.Sanderson. *Proc R Soc* (Lond) 48:89.

Burdon-Sanderson JS, Page FJM. 1879. On the time relations of the excitatory process of the ventricle of the heart of the frog. *J Physiol* (Lond) 2:384.

Caton R. 1875. The electric currents of the brain. *Br Med* 2:278.

DeForest L.U.S. patents 841, 387 (1907) and 879,532 (1908).

Donders FC. 1872. De secondaire contracties order den involed der systolen van het hart, met en zonder vagus-prikkung. Oncerszoek ged in physiol Lab d Utrecht Hoogesch, Bd. 1, S. 256, Derde reeks TI p 246 bis 255.

Einthoven W. 1903. Ein neues Galvanometer. *Ann Phys* 12 (suppl 4): 1059.

Einthoven W, Fahr G, de Waart A. 1913. Uber die Richtung und die manifeste Grosse der Potential schwankunzen in menschlichen Herzen. *Pflugers Arch* 150:275.

Garceau EL, Davis H. 1935. An Ink-writing electroencephalograph. *Arch Neurol Psychiatry* 34:1292.

Gasser HS, Erlanger J. 1922. A study of the action currents of nerve with the cathode ray oscillograph. *Am Physiol* 62:496.
Gasser HS, Newcomer HS. 1921. Physiological action currents in the phrenic nerve. *Am Physiol* 57(1):1.
Geddes LA, Foster KS, Senior J, Kahfeld A. 1989. The Inductorium: The stimulator associated with discovery. *Med Instrum* 23(4):308.
Haynes JR. 1936. Heated stylus for use with waxed recording paper. *Rev Sci Instrum* 7:108.
Hoff HE, Geddes LA. 1957. The rheotome and its prehistory: A study in the historical interrelation of electrophysiology and electromechanics. *Bull Hist Med* 31(3):212.
Kolliker RA, Muller J. 1856. Nachweiss der negativen Schwankung des Muskelstromsnaturlich sic contrakinenden Muskel: Verhandl. *Phys Med Ges Wurzburg* 6:528.
Mann H. 1930–31. A light weight portable EKG. *Am Heart J* 7:796.
Marey EJ. 1885. *Methode Graphique*, 2d ed. Paris, Masson.
Marey EJ. 1876. Des variations electriques des muscles du coeur en particulier etudiee au moyen de l'ectrometre d M. Lippmann. *C R Acad Sci* 82:975.
Marvin HB, et al. 1925. U.S. patent 1,817,913.
Matthews BHC. 1935. Recent developments in electrical instruments for biological and medical purposes. *J Sci Instrum* 12(7):209.
Valentinuzzi ME, Geddes LA, Hoff HE, Bourland JD. 1970. Properties of the 30° hexaxial (EinthovenGoldberger) system of vectorcardiography. *Cardiovasc Res Cent Bull* 9(2):64.
Waller AD. 1887. A demonstration on man of electromotive changes accompanying the heart's beat. *J Physiol* (Lond) 8:229.

III
Medical Instruments and Devices

Wolf W.von Maltzahn
University of Texas at Arlington

N NOT TOO LONG AGO, the term *medical instrument* stood for simple hand-held instruments used by physicians for observing patients, examining organs, making simple measurements, or administering medication. These small instruments, such as stethoscopes, thermometers, tongue depressors, and a few surgical tools, typically fit into a physician's hand bag. Today's medical instruments are considerably more complicated and diverse, primarily because they incorporate electronic systems for sensing, transducing, manipulating, storing, and displaying data or information. Furthermore, med-ical specialists today request detailed and accurate measurements of a vast number of physiologic param-eters for diagnosing illnesses and prescribe complicated procedures for treating these. As a result, the number of medical instruments and devices has grown from a few hundred a generation ago to more than 10,000 today, and the complexity of these instruments has grown at the same pace. The description of all these instruments and devices would fill an entire handbook by itself; however, due to the limited space assigned to this topic, only a selected number are described.

While medical instruments acquire and process information and data for monitoring patients and diagnosing illnesses, medical devices use electrical, mechanical, chemical, or radiation energy for achiev-ing a desired therapeutic purpose, maintaining physiologic functions, or assisting a patient's healing process. To mention only a few functions, medical devices pump blood, remove metabolic waste products, destroy kidney stones, infuse fluids and drugs, stimulate muscles and nerves, cut tissue, administer anesthesia, alleviate pain, restore function, or warm tissue. Because of their complexity, medical devices are used mostly in hospitals and medical centers by trained personnel, but some also can be found in private homes operated by patients themselves or their caregivers.

This section on medical instruments and devices neither replaces a textbook on this subject nor presents the material in a typical textbook manner. The authors assume the reader to be interested in but not knowledgeable on the subject. Therefore, each chapter begins with a short introduction to the subject material, followed by a brief description of current practices and principles, and ends with recent trends and developments. Whenever appropriate, equations, diagrams, and pictures amplify and illustrate the topic, while tables summarize facts and data. The short reference section at the end of each

chapter points toward further resource materials, including books, journal articles, patents, and company brochures.

The chapters in the first half of this section cover the more traditional topics of bioinstrumentation, such as biopotential amplifiers and noninvasive blood pressure, blood flow, and respiration monitors, while those of the second half focus more on recently developed instruments and devices such as pulse oximeters or home-care monitoring devices. Some of this latter material is new or hard to find elsewhere. A few traditional bioinstrumentation or electroencephalography have been omitted entirely because most textbooks on this subject give excellent introductions and reviews. Transducers, biosensors, and electrodes are covered in other sections of this handbook. Thus this section provides an overview, albeit an incom-plete one, of recent developments in the field of medical instruments and devices.

14
Biopotential Amplifiers

Joachim H.Nagel
University of Stuttgart

Biosignals are recorded as potentials, voltages, and electrical field strengths generated by nerves and muscles. The measurements involve voltages at very low levels, typically ranging between 1 μV and 100 mV, with high source impedances and superimposed high level interference signals and noise. The signals need to be amplified to make them compatible with devices such as displays, recorders, or A/D converters for computerized equipment. Amplifiers adequate to measure these signals have to satisfy very specific requirements. They have to provide amplification selective to the physiological signal, reject superimposed noise and interference signals, and guarantee protection from damages through voltage and current surges for both patient and electronic equipment. Amplifiers featuring these specifications are known as *biopotential amplifiers*. Basic requirements and features, as well as some specialized systems, will be presented.

14.1 Basic Amplifier Requirements

The basic requirements that a biopotential amplifier has to satisfy are:

- the physiological process to be monitored should not be influenced in any way by the amplifier
- the measured signal should not be distorted
- the amplifier should provide the best possible separation of signal and interferences
- the amplifier has to offer protection to the patient from any hazard of electrical shock
- the amplifier itself has to be protected against damages that might result from high input voltages as they occur during the application of defibrillators or electrosurgical instrumentation

A typical configuration for the measurement of biopotentials is shown in Fig. 14.1. Three electrodes, two of them picking up the biological signal and the third providing the reference potential, connect the subject to the amplifier. The input signal to the amplifier consists of five components: (1) the desired biopotential, (2) undesired biopotentials, (3) a power line interference signal of 60 Hz (50 Hz in some countries) and its harmonics, (4) interference signals generated by the tissue/electrode interface, and (5) noise. Proper design of the amplifier provides rejection of a large portion of the signal interferences.

0-8493-1813-0/03/$0.00+$ 1.50

The main task of the differential amplifier as shown in Fig. 14.1 is to reject the line frequency interference that is electrostatically or magnetically coupled into the subject. The desired biopotential appears as a

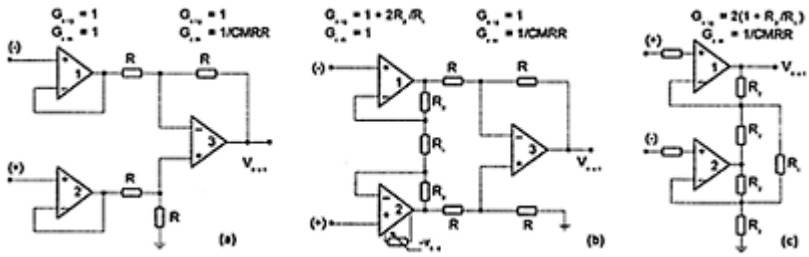


FIGURE 14.1 Typical configuration for the measurement of biopotentials. The biological signal V_{biol} appears between the two measuring electrodes at the right and left arm of the patient, and is fed to the inverting and the noninverting inputs of the differential amplifier. The right leg electrode provides the reference potential for the amplifier with a common mode voltage V_c as indicated.

voltage between the two input terminals of the differential amplifier and is referred to as the *differential signal*. The line frequency interference signal shows only very small differences in amplitude and phase between the two measuring electrodes, causing approximately the same potential at both inputs, and thus appears only between the inputs and ground and is called the *common mode signal* Strong rejection of the common mode signal is one of the most important characteristics of a good biopotential amplifier.

The **common mode rejection ratio** (or CMRR) of an amplifier is defined as the ratio of the differential mode gain over the common mode gain. As seen in Fig. 14.1, the rejection of the common mode signal in a biopotential amplifier is both a function of the amplifier CMRR and the source impedances Z_1 and Z_2. For the ideal biopotential amplifier with Z_1=Z_2 and infinite CMRR of the differential amplifier, the output voltage is the pure biological signal amplified by G_D, the differential mode gain: V_{out}=G_D.V_{biol}. With finite CMRR, the common mode signal is not completely rejected, adding the interference term G_D.V_C/CMRR to the output signal. Even in the case of an ideal differential amplifier with infinite CMRR, the common mode signal will not completely disappear unless the source impedances are equal. The common mode signal V_c causes currents to flow through Z_1 and Z_2. The related voltage drops show a difference if the source impedances are unequal, thus generating a differential signal at the amplifier input

which, of course, is not rejected by the differential amplifier. With amplifier gain G_D and input impedance Z_{in}, the output voltage of the amplifier is:

$$V_{out} = G_D V_{biol} + \frac{G_D V_c}{CMRR} + G_D V_c \left(1 - \frac{Z_{in}}{Z_{in} + Z_1 - Z_2} \right). \qquad (14.1)$$

The output of a real biopotential amplifier will always consist of the desired output component due to a differential biosignal, an undesired component due to incomplete rejection of common mode interference signals as a function of CMRR, and an undesired component due to source impedance unbalance allowing a small proportion of a common mode signal to appear as a differential signal to the amplifier. Since source impedance unbalances of 5000 to 10,000 Ω, mainly caused by electrodes, are not uncommon, and sufficient rejection of line frequency interferences requires a minimum CMRR of 100 dB, the input impedance of the amplifier should be at least 10^9 Ω at 60 Hz to prevent source impedance unbalances from deteriorating the overall CMRR of the amplifier. State-of-the-art biopotential amplifiers provide a CMRR of 120 to 140 dB.

In order to provide optimum signal quality and adequate voltage level for further signal processing, the amplifier has to provide a gain of 100 to 50,000 and needs to maintain the best possible signal-to-noise ratio. The presence of high-level interference signals not only deteriorates the quality of the physiological signals, but also restricts the design of the biopotential amplifier. Electrode half-cell potentials, for example, limit the gain factor of the first amplifier stage since their amplitude can be several

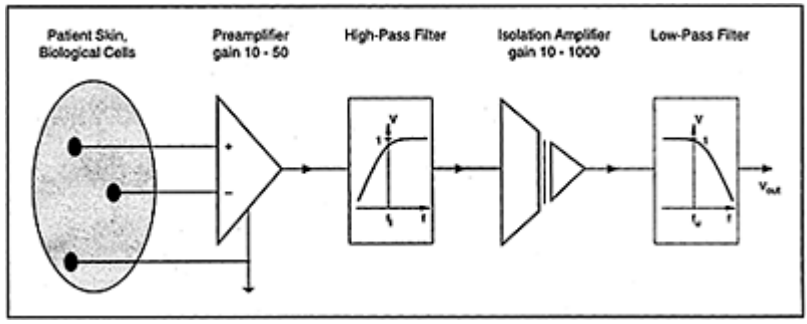


FIGURE 14.2 Schematic design of the main stages of a biopotential amplifier. Three electrodes connect the patient to a preamplifier stage. After removing dc and low-frequency interferences, the signal is connected to an output low-pass filter through an isolation stage that provides electrical safety to the patient, prevents ground loops, and

reduces the influence of interference signals.

orders of magnitude larger than the amplitude of the physiological signal. To prevent the amplifier from going into saturation, this component has to be eliminated before the required gain can be provided for the physiological signal.

A typical design of the various stages of a biopotential amplifier is shown in Fig. 14.2. The electrodes that provide the transition between the ionic flow of currents in biological tissue and the electronic flow of current in the amplifier, represent a complex electrochemical system that is described elsewhere in this handbook. The electrodes determine to a large extent the composition of the measured signal. The preamplifier represents the most critical part of the amplifier itself since it sets the stage for the quality of the biosignal. With proper design, the preamplifier can eliminate, or at least minimize, most of the signals interfering with the measurement of biopotentials.

In addition to electrode potentials and electromagnetic interferences, noise—generated by the amplifier and the connection between biological source and amplifier—has to be taken into account when designing the preamplifier. The total source resistance R_s, including the resistance of the biological source and all transition resistances between signal source and amplifier input, causes thermal voltage noise with a root mean square (rms) value of:

$$E_{rms} = \sqrt{4kTR_sB} \quad (Volt), \tag{14.2}$$

where k=Boltzmann constant, T=absolute temperature, R_s=resistance in Ω, and B=bandwidth in Hz.

Additionally, there is the inherent amplifier noise. It consists of two frequency-dependent components, the internal voltage noise source e_n and the voltage drop across the source resistance R_s caused by an internal current noise generator i_n. The total input noise for the amplifier with a bandwidth of $B=f_2-f_1$ is calculated as the sum of its three independent components:

$$E_{rms}^2 = \int_{f_1}^{f_2} e_n^2 df + R_s^2 \int_{f_1}^{f_2} i_n^2 df + 4kTR_sB\,. \tag{14.3}$$

High signal-to-noise ratios thus require the use of very low noise amplifiers and the limitation of bandwidth. Current technology offers differential amplifiers with voltage noise of less than $10\ \mathrm{nV}/\sqrt{Hz}$ and current noise less than $1\ \mathrm{pA}/\sqrt{Hz}$ Both parameters are frequency dependent and decrease approx-imately with the square root of frequency. The exact relationship depends on the technology of the amplifier input stage. Field effect transistor (FET) preamplifiers exhibit about five times the voltage noise density compared to bipolar transistors but a current noise density that is about 100 times smaller.

The purpose of the high-pass and low-pass filters in Fig. 14.2 is to eliminate interference signals like electrode half-cell potentials and preamplifier offset potentials and to reduce the noise amplitude by the limitation of the amplifier bandwidth. Since the

biosignal should not be distorted or attenuated, higher-order sharp-cutting linear phase filters have to be used. Active Bessel filters are preferred filter types due to their smooth transfer function. Separation of biosignal and interference is in most cases incomplete due to the overlap of their spectra.

The isolation stage serves the galvanic decoupling of the patient from the measuring equipment and provides safety from electrical hazards. This stage also prevents galvanic currents from deteriorating the signal-to-noise ratio especially by preventing ground loops. Various principles can be used to realize the isolation stage. Analog isolation amplifiers use either transformer, optical, or capacitive couplers to transmit the signal through the isolation barrier. Digital isolation amplifiers use a voltage/frequency converter to digitize the signal before it is transmitted easily by optical or inductive couplers to the output frequency/ voltage converter. The most important characteristics of an isolation amplifier are low leakage current, isolation impedance, isolation voltage (or mode) rejection (IMR), and maximum safe isolation voltage.

Interferences

The most critical point in the measurement of biopotentials is the contact between electrodes and biological tissue. Both the electrode offset potential and the electrode/tissue impedance are subject to changes due to relative movements of electrode and tissue. Thus, two interference signals are generated as motion artifacts: the changes of the electrode potential and motion-induced changes of the voltage drop caused by the input current of the preamplifier. These motion artifacts can be minimized by providing high input impedances for the preamplifier, usage of nonpolarized electrodes with low half-cell potentials such as Ag/AgCl electrodes, and by reducing the source impedance by use of electrode gel. Motion artifacts, interferences from external electromagnetic fields, and noise can also be generated in the wires connecting electrodes and amplifier. Reduction of these interferences is achieved by using twisted pair cables, shielded wires, and *input guarding*.

Recording of biopotentials is often done in an environment that is equipped with many electrical systems that produce strong electrical and magnetic fields. In addition to 60-Hz power line frequency and some strong harmonics, high-frequency electromagnetic fields are encountered. At power line frequency, the electric and magnetic components of the interfering fields can be considered separately. Electrical fields are caused by all conductors that are connected to power, even with no flow of current. A current is capacitively coupled into the body where it flows to the ground electrode. If an isolation amplifier is used without patient ground, the current is capacitively coupled to ground. In this case, the body potential floats with a voltage of up to 100 V toward ground. Minimizing interferences requires increasing the distance between power lines and the body, use of isolation amplifiers, separate grounding of the body at a location as far away from the measuring electrodes as possible, and use of shielded electrode cables.

The magnetic field components produce eddy currents in the body. The amplifier, the electrode cable, and the body form an induction loop that is subject to the generation of an interference signal. Minimizing this interference signal requires increasing the distance between the interference source and patient, twisting the connecting cables, shielding of the magnetic fields, and relocating the patient to a place and orientation that

offers minimum interference signals. In many cases, an additional narrow band-rejection filter (notch filter) is implemented as an additional stage in the biopotential amplifier to provide sufficient suppression of line frequency interferences.

In order to achieve optimum signal quality, the biopotential amplifier has to be adapted to the specific application. Based on the signal parameters, both appropriate bandwidth and gain factor are chosen. Figure 14.3 shows an overview of the most commonly measured biopotentials and specifies the normal ranges for amplitude and bandwidth.

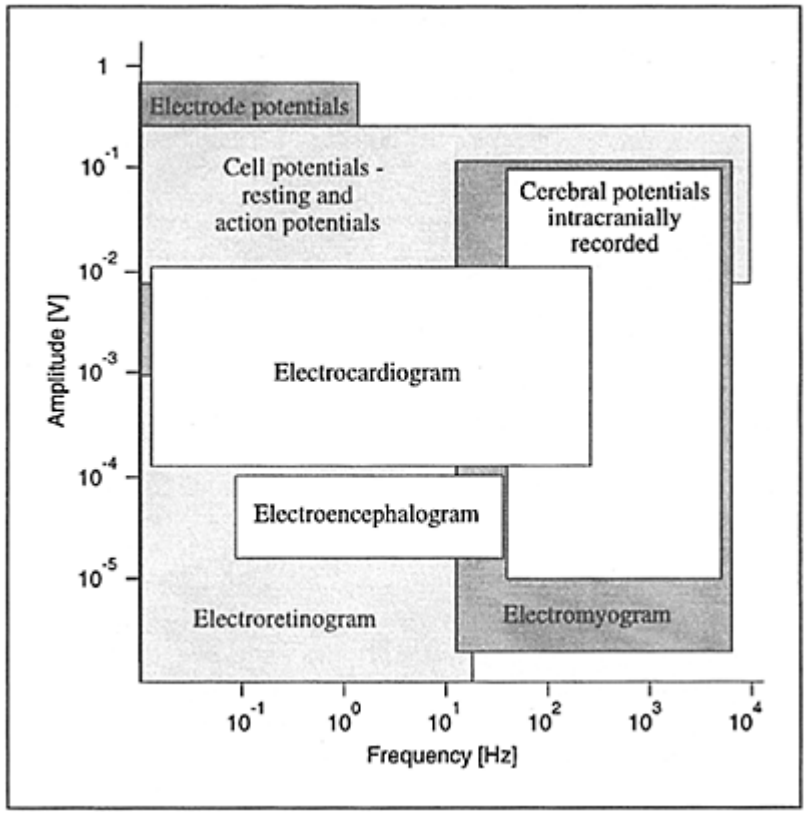


FIGURE 14.3 Amplitudes and spectral ranges of some important biosignals. The various biopotentials completely cover the area from 10^{-6} V to almost 1 V and from dc to 10 kHz.

A final requirement for biopotential amplifiers is the need for calibration. Since the amplitude of the biopotential often has to be determined very accurately, there must be provisions to easily determine the gain or the amplitude range referenced to the input of

the amplifier. For this purpose, the gain of the amplifier must be well calibrated. In order to prevent difficulties with calibrations, some amplifiers that need to have adjustable gain use a number of fixed gain settings rather than providing a continuous gain control. Some amplifiers have a standard signal source of known amplitude built in that can be momentarily connected to the input by the push of a button to check the calibration at the output of the biopotential amplifier.

14.2 Special Circuits

14.2.1 Instrumentation Amplifier

An important stage of all biopotential amplifiers is the input preamplifier that substantially contributes to the overall quality of the system. The main tasks of the preamplifier are to sense the voltage between two measuring electrodes while rejecting the common mode signal, and minimizing the effect of electrode polarization overpotentials. Crucial to the performance of the preamplifier is the input impedance, which should be as high as possible. Such a differential amplifier cannot be realized using a standard single **operational amplifier** (op-amp) design since this does not provide the necessary high input impedance. The general solution to the problem involves voltage followers, or noninverting amplifiers, to attain high input impedances. A possible realization is shown in Fig. 14.4a. The main disadvantage of this circuit is that it requires high CMRR both in the followers and in the final op-amp. With the input buffers working at unity gain, all the common-mode rejection must be accomplished in the output amplifier, requiring very

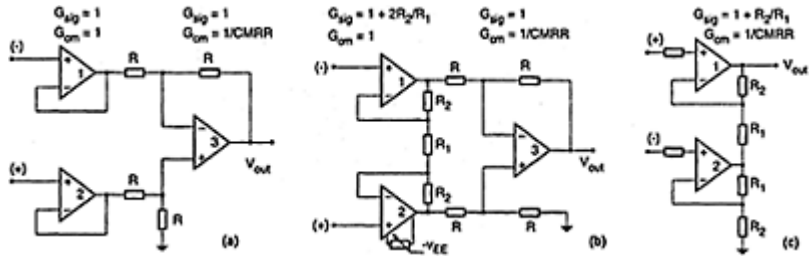


FIGURE 14.4 Circuit drawings for three different realizations of instrumentation amplifiers for biomedical applications. Voltage follower input stage (a), improved, amplifying input stage (b), and 2-op-amp version (c).

precise resistor matching. Additionally, the noise of the final op-amp is added at a low signal level, decreasing the signal-to-noise ratio unnecessarily. The circuit in Fig. 14.4b eliminates this disadvantage. It represents the standard instrumentation amplifier configuration. The two input op-amps provide high differential gain and unity common-mode gain without the requirement of close resistor matching. The differential output from the first stage represents a signal with substantial relative reduction of the common-mode signal and is used to drive a standard differential amplifier that further reduces the common-mode signal. CMRR of the output op-amp as well as resistor matching in its circuit are less critical than in the follower type instrumentation amplifier. Offset trimming for the whole circuit can be done at one of the input op-amps. Complete instrumentation amplifier integrated circuits based on this standard instrumentation amplifier configuration are available from several manufacturers. All components except R_1, which determines the gain of the amplifier, and the potentiometer for offset trimming are contained on the integrated circuit chip. Figure 14.4c shows another configuration that offers high input impedance with only two op-amps. For good CMRR, however, it requires precise resistor matching.

In applications where DC and very low-frequency biopotentials are not to be measured, it would be desirable to block those signal components at the preamplifier inputs by simply adding a capacitor working as a passive high-pass filter. This would eliminate the electrode offset potentials and permit a higher gain factor for the preamplifier *and thus a higher CMRR*. A capacitor between electrodes and amplifier input would, however, result in charging effects from the input bias current. Due to the difficulty of precisely matching capacitors for the two input leads, they would also contribute to an increased source impedance unbalance and thus reduce CMRR. Avoiding the problem of charging effects by adding a resistor between the preamplifier inputs and ground as shown in Fig. 14.5a also results in a decrease of CMRR due to the diminished and mismatched input impedance. A 1% mismatch for two 1-MΩ, resistors can already create a −60-dB loss in CMRR. The loss in CMRR is much greater if the capacitors are mismatched, which cannot

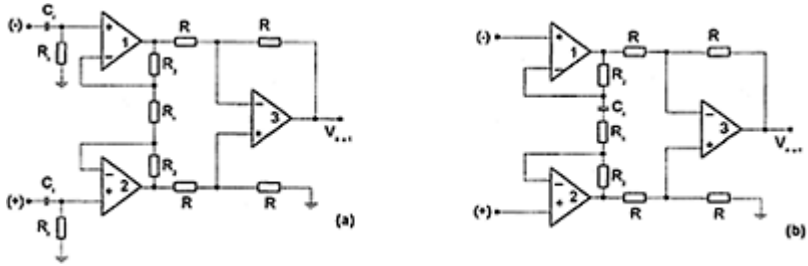


FIGURE 14.5 AC coupled instrumentation amplifier designs. The classical design using an RC high-pass filter at the inputs (a), and a high CMRR "quasi-high-pass" amplifier as proposed by Lu (b).

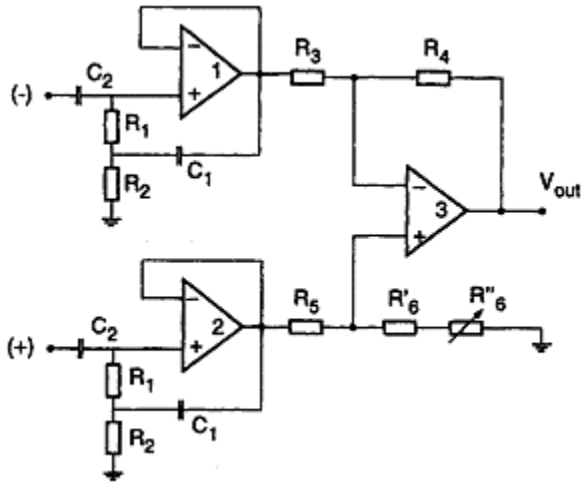


FIGURE 14.6 Composite instrumentation amplifier based on an ac-coupled first stage. The second stage is based on a one op-amp differential amplifier that can be replaced by an instrumentation amplifier.

be prevented in real systems. Nevertheless, such realizations are used where the specific situation allows. In some applications, a further reduction of the amplifier to a two-electrode amplifier configuration would be convenient, even at the expense of some loss in the CMRR. Figure 14.6 shows a preamplifier design working with two electrodes and providing AC coupling as proposed by Pallás-Areny [1990].

A third alternative of eliminating DC and low frequencies in the first amplifier stage is a directly coupled quasi-high-pass amplifier design, which maintains the high CMRR of DC coupled high input impedance instrumentation amplifiers [Song et al., 1998]. In this design, the gain determining resistor R_1 (Fig. 14.5a) is replaced by a first-order, high-pass filter consisting of R_1 and a series capacitor C_f. The signal gain of the amplifier is

$$G = 1 + \frac{2R_2}{R_1 + \dfrac{1}{j\omega C}}. \qquad (14.4)$$

Thus, DC gain is 1, while the high-frequency gain remains at $G=1+2R_2/R_1$. A realization using an offthe-shelf instrumentation amplifier (Burr-Brown INA 118) operates at low power (0,35 mA) with low-offset voltage (11 μV typical) and low-input bias current (1 nA typical), and offers a high CMRR of 118 dB at a gain of G=50. The very high-input impedance (10 GΩ) of the instrumentation amplifier renders it insensitive to fluctuations

of the electrode impedance. Therefore, it is suitable for bioelectric measurements using pasteless electrodes applied to unprepared, i.e., high impedance skin.

The preamplifier, often implemented as a separate device that is placed close to the electrodes or even directly attached to the electrodes, also acts as an impedance converter that allows the transmission of even weak signals to the remote monitoring unit. Due to the low-output impedance of the preamplifier, the input impedance of the following amplifier stage can be low, and still the influence of interference signals coupled into the transmission lines is reduced.

Isolation Amplifier and Patient Safety

Isolation amplifiers can be used to break ground loops, eliminate source ground connections, and provide isolation protection to the patient and electronic equipment. In a biopotential amplifier, the main purpose of the isolation amplifier is the protection of the patient by eliminating the hazard of electric shock resulting from the interaction among the patient, amplifier, and other electric devices in the patient's environment, specifically defibrillators and electrosurgical equipment. It also adds to the prevention of line frequency interferences.

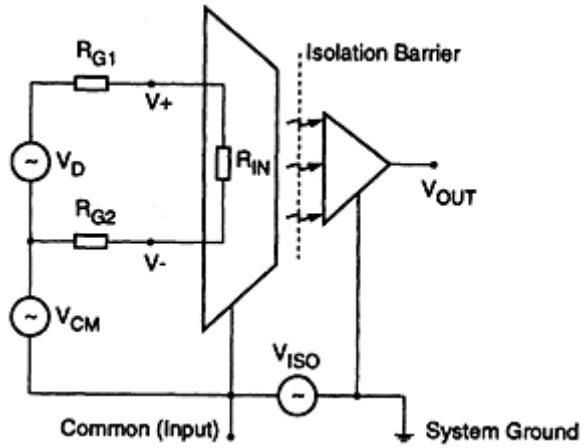


FIGURE 14.7 Equivalent circuit of an isolation amplifier. The differential amplifier on the left transmits the signal through the isolation barrier by a transformer, capacitor, or an opto-coupler.

Isolation amplifiers are realized in three different technologies: transformer isolation, capacitor isolation, and opto-isolation. An isolation barrier provides a complete galvanic separation of the input side, i.e., patient and preamplifier, from all equipment on the output side. Ideally, there will be no flow of electric current across the barrier. The

isolation-mode voltage is the voltage that appears across the isolation barrier, i.e., between the input common and the output common (Fig. 14.7). The amplifier has to withstand the largest expected isolation voltages witnout damage. Two isolation voltages are specified for commercial isolation amplifiers: (1) the continuous rating and (2) the test voltage. To eliminate the need for longtime testing, the device is tested at about two times the rated continuous voltage. Thus, for a continuous rating of 2000 V, the device has to be tested at 4000 to 5000 V for a reasonable period of time.

Since there is always some leakage across the isolation barrier, the **isolation mode rejection ratio** (IMRR) is not infinite. For a circuit as shown in Fig. 14.7, the output voltage is:

$$V_{out} = \frac{G}{R_{G1} + R_{G2} + R_{IN}} \left[V_D + \frac{V_{CM}}{CMRR} \right] + \frac{V_{ISO}}{IMRR}, \qquad (14.5)$$

where G is the amplifier gain, V_D, V_{CM}, and V_{ISO} are differential, common mode, and isolation voltages, respectively, and CMRR is the common mode rejection ratio for the amplifier [Burr-Brown, 1994].

Typical values of IMRR for a gain of 10 are 140 dB at DC, and 120 dB at 60 Hz with a source unbalance of 5000 Ω. The isolation impedance is approximately 1.8 pF || 10^{12} Ω.

Transformer coupled isolation amplifiers perform on the basis of inductive transmission of a carrier signal that is amplitude modulated by the biosignal. A synchronous demodulator on the output port reconstructs the signal before it is fed through a Bessel response low-pass filter to an output buffer. A power transformer, generally driven by a 400 to 900 kHz square wave, supplies isolated power to the amplifier.

Optically coupled isolation amplifiers can principally be realized using only a single LED and photodiode combination. While useful for a wide range of digital applications, this design has fundamental limitations as to its linearity and stability as a function of time and temperature. A matched photodiode design, as used in the Burr-Brown 3650/3652 isolation amplifier, overcomes these difficulties [Burr-Brown, 1994]. Operation of the amplifier requires an isolated power supply to drive the input stages. Transformer coupled low leakage current isolated DC/DC converters are commonly used for this purpose. In some particular applications, especially in cases where the signal is transmitted over a longer distance by fiber optics, e.g., ECG amplifiers used for gated magnetic resonance imaging, batteries are used to power the amplifier. Fiber-optic coupling in isolation amplifiers is another option that offers the advantage of higher flexibility in the placement of parts on the amplifier board.

Biopotential amplifiers have to provide sufficient protection from electrical shock to both user and patient. Electrical-safety codes and standards specify the minimum safety requirements for the equipment, especially the maximum leakage currents for chassis and patient leads, and the power distribution system [Webster, 1992; AAMI, 1993].

Special attention to patient safety is required in situations where biopotential amplifiers are connected to personal computers that are more and more often used to process and store physiological signals and data. Due to the design of the power supplies used in standard PCs permitting high leakage currents—an inadequate situation for a medical environment—there is a potential risk involved even when the patient is isolated

from the PC through an isolation amplifier stage or optical signal transmission from the amplifier to the computer. This holds especially in those cases where, due to the proximity of the PC to the patient, an operator might touch the patient and the computer at the same time, or the patient might touch the computer. It is required that a special power supply with sufficient limitation of leakage currents is used in the computer, or that an additional, medical-grade isolation transformer is used to provide the necessary isolation between power outlet and PC.

Surge Protection

The isolation amplifiers described in the preceding paragraph are primarily used for the protection of the patient from electric shock. Voltage surges between electrodes as they occur during the application of a defibrillator or electrosurgical instrumentation also present a risk to the biopotential amplifier. Biopotential amplifiers should be protected against serious damage to the electronic circuits. This is also part of patient safety since defective input stages could otherwise apply dangerous current levels to the patient. To achieve this protection, voltage-limiting devices are connected between each measuring electrode and electric ground. Ideally, these devices do not represent a shunt impedance and thus do not lower the input impedance of the preamplifier as long as the input voltage remains in a range considered safe for the equipment. They appear as an open circuit. As soon as the voltage drop across the device reaches a critical value V_b, the impedance of the device changes sharply and current passes through it to such an extent that the voltage cannot exceed V_b due to the voltage drop across the series resistor R as indicated in Fig. 14.8.

Devices used for amplifier protection are diodes, Zener diodes, and gas-discharge tubes. Parallel silicon diodes limit the voltage to approximately 600 mV. The transition from nonconducting to conducting state is not very sharp, and signal distortion begins at about 300 mV, which can be within the range of input voltages depending on the electrodes used. The breakdown voltage can be increased by connecting several diodes in series. Higher breakdown voltages are achieved by Zener diodes connected back to back. One of

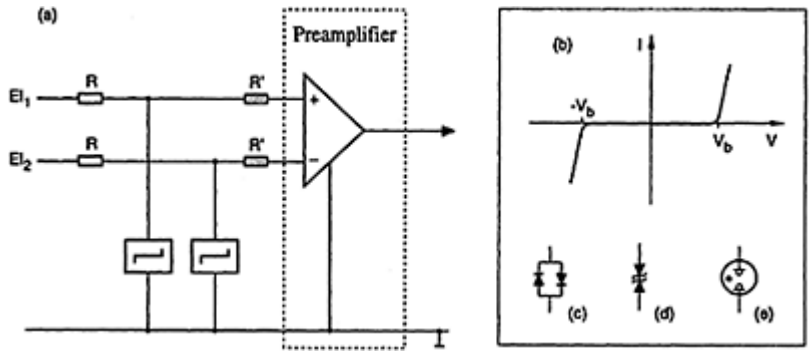


FIGURE 14.8 Protection of the amplifier input against high-voltage

transients. The connection diagram for voltage-limiting elements is shown in panel (a) with two optional resistors R′ at the input. A typical current-voltage characteristic is shown in panel (b). Voltage-limiting elements shown are the antiparallel connection of diodes (c), antiparallel connection of Zener diodes (d), and gas-discharge tubes (e).

the diodes will be biased in the forward direction and the other in the reverse direction. The breakdown voltage in the forward direction is approximately 600 mV, but the breakdown voltage in the reverse direction is higher, generally in the range of 3 to 20 V, with a sharper voltage-current characteristic than the diode circuit.

A preferred voltage-limiting device for biopotential amplifiers is the *gas-discharge tube*. Due to its extremely high impedance in the nonconducting state, this device appears as an open circuit until it reaches its breakdown voltage. At the breakdown voltage, which is in the range of 50 to 90 V, the tube switches to the conducting state and maintains a voltage that is usually several volts less than the breakdown voltage. Though the voltage maintained by the gas-discharge tube is still too high for some amplifiers, it is low enough to allow the input current to be easily limited to a safe value by simple circuit elements such as resistors like the resistors R′ indicated in Fig. 14.8a. Preferred gas discharge tubes for biomedical applications are miniature neon lamps that are very inexpensive and have a symmetric characteristic.

Input Guarding

The common mode input impedance and thus the CMRR of an amplifier can be greatly increased by guarding the input circuit. The common mode signal can be obtained by two averaging resistors connected between the outputs of the two input op-amps of an instrumentation amplifier as shown in Fig. 14.9. The buffered common-mode signal at the output of op-amp 4 can be used as guard voltage to reduce the effects of cable capacitance and leakage.

In many modern biopotential amplifiers, the reference electrode is not grounded. Instead, it is connected to the output of an amplifier for the common mode voltage, op-amp 3 in Fig. 14.10, which works

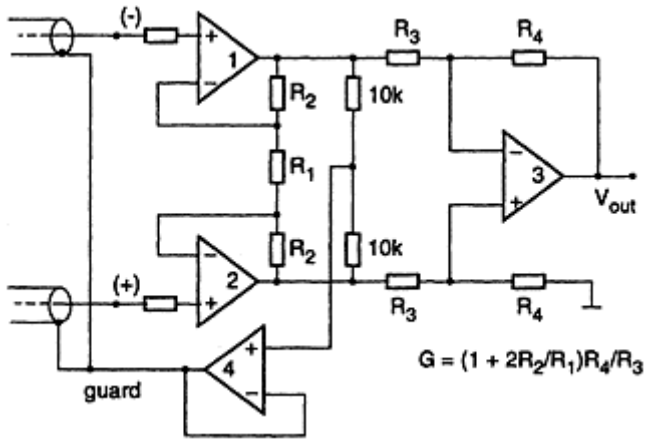


FIGURE 14.9 Instrumentation amplifier providing input guarding.

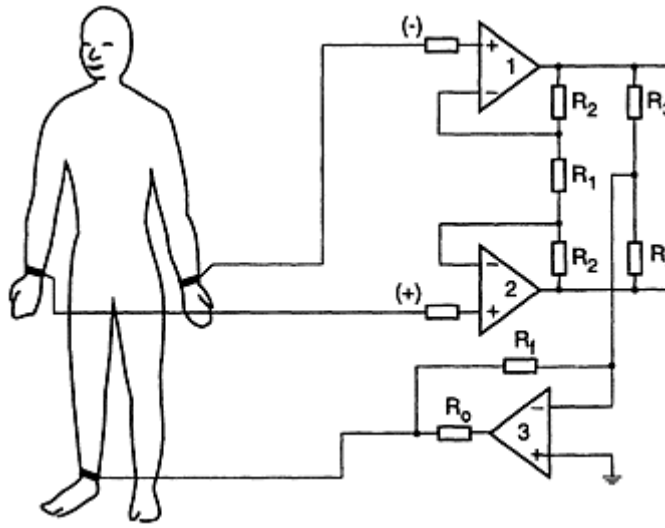


FIGURE 14.10 Driven-right-leg circuit reducing common-mode interference.

as an inverting amplifier. The inverted common mode voltage is fed back to the reference electrode. This negative feedback reduces the common-mode voltage to a low value [Webster, 1992]. Electrocardiographs based on this principle are called driven-right-leg systems replacing the right leg ground electrode of ordinary electrocardiographs by an actively driven electrode.

Dynamic Range and Recovery

With an increase of either the common mode or differential input voltage, there will be a point where the amplifier will overload and the output voltage will no longer be representative for the input voltage. Similarly, with a decrease of the input voltage, there will be a point where the noise components of the output voltage cover the output signal to a degree that a measurement of the desired biopotential is no longer possible. The dynamic range of the amplifier, i.e., the range between the smallest and largest possible input signal to be measured, has to cover the whole amplitude range of the physiological signal of interest. The required dynamic range of biopotential amplifiers can be quite large. In an application like fetal monitoring for example, two signals are recorded simultaneously from the electrodes that are quite different in their amplitudes: the fetal and the maternal EGG. While the maternal EGG shows an amplitude of up to 10 mV, the fetal EGG often does not reach more than 1 µV. Assuming that the fetal EGG is separated from the composite signal and fed to an analog/digital converter for digital signal processing with a resolution of 10 bit (signed integer), the smallest voltage to be safely measured with the biopotential amplifier is 1/512 µV or about 2 nV vs. 10 mV for the largest signal, or even up to 300 mV in the presence of an electrode offset potential. This translates to a dynamic range of 134 dB for the signals alone and 164 dB if the electrode potential is included in the consideration. Though most applications are less demanding, even such extreme requirements can be realized through careful design of the biopotential amplifer and the use of adequate components. The penalty for using less expensive amplifiers with diminished performance would be a potentially severe loss of information.

Transients appearing at the input of the biopotential amplifier, like voltage peaks from a cardiac pacemaker or a defibrillator, can drive the amplifier into saturation. An important characteristic for the amplifier is the time it takes to recover from such overloads. The recovery time depends on the characteristics of the transient, like amplitude and duration, the specific design of the amplifier, like bandwidth, and the components used. Typical biopotential amplifiers may take several seconds to recover from severe overload. The recovery time can be reduced by disconnecting the amplifier inputs at the discovery of a transient using an electronic switch.

Passive Isolation Amplifiers

Increasingly, biopotentials have to be measured within implanted devices and need to be transmitted to an external monitor or controller. Such applications include cardiac pacemakers transmitting the intracardiac EGG and functional electrical stimulation where, e.g., action potentials measured at one eyelid serve to stimulate the other lid to restore the physiological function of a damaged lid at least to some degree. In these applications, the power consumption of the implanted biopotential amplifier limits the lifespan of the implanted device. The usual solution to this problem is an inductive transmission of power into the implanted device that serves to recharge an implanted battery. In applications where the size of the implant is of concern, it is desirable to eliminate the need for the battery and the related circuitry by using a quasipassive biopotential amplifer, i.e., an amplifier that does not need a power supply.

The function of passive telemetric amplifiers for biopotentials is based on the ability of the biological source to drive a low power device such as a FET and the sensing of the biopotentials through inductive or acoustic coupling of the implanted and external devices [Nagel, 1982]. In an inductive system, a FET serves as a load to an implanted secondary LC-circuit which is stimulated inductively by an extracorporal oscillator (Fig. 14.11). Depending on the special realization of the system, the biopotential is available in the external circuit from either an amplitude or frequency-modulated carrier-signal. The input impedance of the inductive transmitter as a function of the secondary load impedance Z_2 is given by:

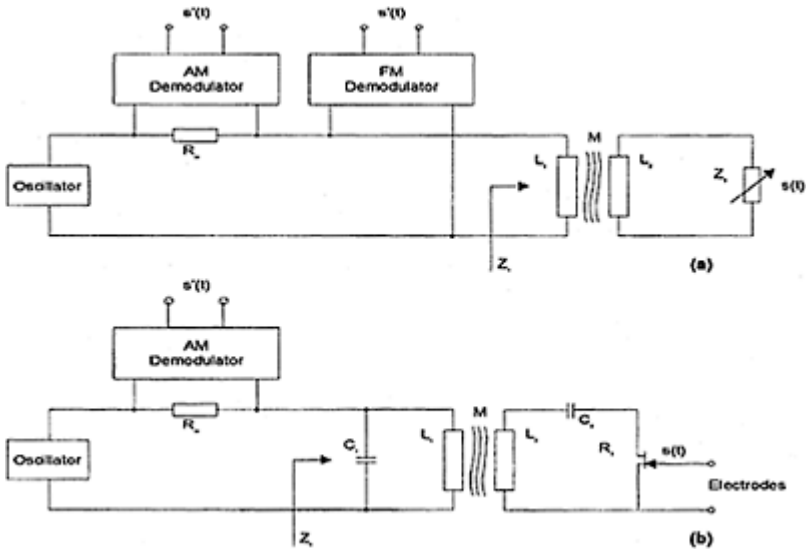


FIGURE 14.11 The *passive* isolation amplifier in (a) can be operated without the need for an isolated power supply. The biological source provides the power to modulate the load impedance of an inductive transformer. As an easy realization shown in panel (b), a FET can be directly connected to two electrodes. The source-drain resistance changes as a linear function of the biopotential which is then reflected by the input impedance of the transformer.

$$Z_1 = j\omega L_1 + \frac{(\omega M)^2}{Z_2 + j\omega L_2}. \qquad (14.6)$$

In an amplitude-modulated system, the resistive part of the input-impedance Z_1 must change as a linear function of the biopotential. The signal is obtained as the envelope of the carrier signal, measured across a resistor R_m. A frequency-modulated system is realized when the frequency of the signal generator is determined at least in part by the impedance Z_1 of the inductive transmitter. In both cases, the signal-dependent changes of the secondary impedance Z_2 can be achieved by a junction-FET. Using the field effect transistor as a variable load resistance changing its resistance in proportion to the source-gate voltage, which is determined by the electrodes of this two-electrode amplifier, the power supplied by the biological source is sufficient to drive the amplifier. The input impedance can be in the range of $10^{10}\,\Omega$.

Optimal transmission characteristics are achieved with AM systems. Different combinations of external and implanted resonance circuits are possible to realize in an AM system, but primary parallel with secondary serial resonance yields the best characteristics. In this case, the input impedance is given by:

$$Z_1 = \frac{1}{j\omega C_1} + \left(\frac{L_1}{M}\right)^2 \cdot R_2. \qquad (14.7)$$

The transmission factor $(L_1/M)^2$ is optimal since the secondary inductivity, i.e., the implanted inductivity, can be small, only the external inductivity determines the transmission factor and the mutual inductivity should be small, a fact that favors the loose coupling that is inherent to two coils separated by skin and tissue. There are, of course, limits to M which cannot be seen from Eq. (14.7). In a similar fashion, two piezoelectric crystals can be employed to provide the coupling between input and output.

This 2-lead isolation amplifier design is not limited to telemetric applications. It can also be used in all other applications where its main advantage lies in its simplicity and the resulting substantial cost savings as compared to other isolation amplifiers that require additional amplifier stages and an additional isolated power supply.

Digital Electronics

The ever-increasing density of integrated digital circuits together with their extremely low power consumption permits digitizing and preprocessing of signals already on the isolated patient-side of the amplifiers, thus improving signal quality and eliminating the problems normally related to the isolation barrier, especially those concerning isolation voltage interferences and long-term stability of the isolation amplifiers. Digital signal transmission to a remote monitoring unit, a computer system, or computer network can be achieved without any risk of picking up transmission line interferences, especially when implemented with fiberoptical cables.

Digital techniques also offer an easy means of controlling the front-end of the amplifier. Gain factors can be easily adapted, and changes of the electrode potential

resulting from electrode polarization or from interferences that might drive the differential amplifier into saturation can easily be detected and compensated.

14.3 Summary

Biopotential amplifiers are a crucial component in many medical and biological measurements, and largely determine the quality and information content of the measured signals. The extremely wide range of necessary specifications with regard to bandwidth, sensitivity, dynamic range, gain, CMRR, and patient safety leaves only little room for the application of general-purpose biopotential amplifiers, and mostly requires the use of special purpose amplifiers.

Defining Terms

Common Mode Rejection Ratio (CMRR): The ratio between the amplitude of a common mode signal and the amplitude of a differential signal that would produce the same output amplitude or as the ratio of the differential gain over the common-mode gain: $CMRR=G_D/G_{CM}$. Expressed in decibels, the common mode rejection is $20 \log_{10}$ CMRR. The common mode rejection is a function of frequency and source-impedance unbalance.

Isolation Mode Rejection Ratio (IMRR): The ratio between the isolation voltage, V_{ISO}, and the amplitude of the isolation signal appearing at the output of the isolation amplifier, or as isolation voltage divided by output voltage V_{OUT} in the absence of differential and common mode signal: $IMRR = V_{ISO}/V_{OUT}$.

Operational Amplifier (op-amp): A very high-gain de-coupled differential amplifier with single-ended output, high-voltage gain, high-input impedance, and low-output impedance. Due to its high open-loop gain, the characteristics of an op-amp circuit only depend on its feedback network. Therefore, the integrated circuit op-amp is an extremely convenient tool for the realization of linear amplifier circuits [Horowitz and Hill, 1980].

References

AAMI, 1993. *AAMI Standards and Recommended Practices, Biomedical Equipment,* Vol. 2, 4th ed. AAMI, Arlington, VA.

Burr-Brown, 1994. *Burr-Brown Integrated Circuits Data Book, Linear Products,* Burr-Brown Corp., Tucson, AZ.

Horowitz, P. and Hill, W., 1980. *The Art of Electronics,* Cambridge University Press, Cambridge, UK.

Hutten, H., Hrsg., 1992. *Biomedizinische Technik,* Springer-Verlag, Berlin.

Nagel, J., Ostgen, M., and Schaldach, M., *Telemetriesystem,* German Patent Application, *P 3233240.8–15,* 1982.

Pallás-Areny, R. and Webster, J.G. 1990. Composite Instrumentation Amplifier for Biopotentials. *Annals of Biomedical Engineering.* 18, 251–262.

Strong, P., 1970. *Biophysical Measurements,* Tektronix, Inc., Beaverton, OR.

Webster, J.G., Ed., 1992. *Medical Instrumentation, Application and Design*, 2nd ed. Houghton Mifflin Company, Boston, MA.

Song, Y., Ozdamar, O., and Lu, C.C., 1998. Pasteless Electrode/Amplifier System for Auditory Brainstem Response (ABR) Recording. *Annals of Biomedical Engineering.* 26, S-103.

Further Information

Detailed information on the realization of amplifiers for biomedical instrumentation and the availability of commercial products can be found in the references and in the Data Books and Application Notes of various manufacturers of integrated circuit amplifiers like Burr-Brown, Analog Devices, and Precision Monolithics Inc. as well as manufacturers of laboratory equipment like Gould and Grass.

15
Noninvasive Arterial Blood Pressure and Mechanics

Gary Drzewiecki
Rutgers State University

15.1 Introduction

The beginnings of noninvasive arterial pulse recording can be traced to the Renaissance. At that time, the Polish scientist, Strus [1555], had proposed that the arterial pulse possesses a waveform. Although instrumentation that he used was simple, he suggested that changes in the arterial pulse shape and strength may be related to disease conditions. Today, even though the technology is more advanced, noninvasive arterial blood pressure measurement still remains a challenge. Rigorous methods for extracting functional cardiovascular information from noninvasive pressure have been limited.

In this chapter, the most standard of noninvasive methods for arterial pressure measurements will be reviewed and future trends will be proposed. Two types of methods for noninvasive arterial pressure measurement may be defined; those that periodically sample and those that continuously record the pulse waveform. The sampling methods typically provide systolic and diastolic pressure and sometimes mean pressure. These values are collected over different heartbeats during the course of 1 min. The continuous recording methods provide beat-to-beat measurements and, often, the entire waveform. Some continuous methods provide only pulse pressure waveform and timing information.

The knowledge of systolic and diastolic pressure is fundamental for the evaluation of basic cardiovascular function and identifying disease. The choice of method depends on the type of study. For example, high blood pressure is a known precursor to many other forms of cardiovascular disease. A noninvasive method that samples blood pressure over the course of months is usually adequate to study the progression of hypertension. The occlusive cuff-based methods fall into this category. These methods have been automated with recent instruments designed for ambulatory use [Graettinger et al., 1988]. Twenty-four-to 48-h ambulatory monitors have been applied to monitor the diurnal variation of a patient's blood pressure. This type of monitoring can alleviate the problem of "white-coat hypertension," i.e., the elevation of blood pressure associated with a visit to a physician's office [Pickering et al., 1988].

0–8493–1813–0/03/$0.00+$ 1.50

Short-term hemodynamic information obtained from the noninvasive arterial pulse waveform is a virtually untapped arena. While a great deal of knowledge has been gathered on the physics of the arterial pulse [Noordergraaf, 1978], it has been lacking in application because continuous pressure waveform monitors have not been available. A review of methods for continuous pulse monitoring that fill this gap will be provided.

Some applications for pulse monitoring can be found. In the recording time span of less than 1 min, the importance of the pressure waveform dominates, as well as beat-to-beat variations. This type of monitoring is critical in situations where blood pressure can alter quickly, such as due to trauma or anesthesia. Other applications for acute monitoring have been in aerospace, biofeedback, and lie detection. Moreover, pulse dynamics information becomes available such as wave reflection [Li, 1986]. Kelly et al. [1989] have shown that elevated systolic pressure can be attributed to pulse waveform changes due to a decrease in arterial compliance and increased wave reflection. Lastly, information on the cardiovascular control process can be obtained from pulse pressure variability [Omboni et al., 1993].

It has become increasingly popular to provide simultaneous recording of such variables as noninvasive blood pressure, oxygen saturation via pulse oximetry, body temperature, etc., in a single instrument. It is apparent that the advances in computer technology impinge on this practice, making it a clear trend. While this practice is likely to continue, the forefront of this approach will be those instruments that provide more than just a mere marriage of technology in a single design. It will be possible to extract functional information in addition to just pressure. As an example of this, a method for noninvasive measurement of the arterial pressure-lumen area curve will be provided.

15.2 Long-Term Sampling Methods

15.2.1 Vascular Unloading Principle

The vascular unloading principle is fundamental to all occlusive cuff-based methods of determining arterial blood pressure. It is performed by applying an external compression pressure or force to a limb such that it is transmitted to the underlying vessels. It is usually assumed that the external pressure and the tissue pressure (stress) are in equilibrium. The underlying vessels are then subjected to altered transmural pressure (internal minus external pressure) by varying the external pressure. It is further assumed [Marey, 1885] that the tension within the wall of the vessel is zero when transmural pressure is zero. Hence, the term vascular unloading originated.

Various techniques that attempt to detect vascular unloading have been developed. These generally rely on the fact that once a vessel is unloaded, further external pressure will cause it to collapse. In summary,

$$\text{If } P_a > P_c \Rightarrow \text{Lumen open}$$

or

$$\text{if } P_a < P_c \Rightarrow \text{Lumen closed}$$

where P_a is the arterial pressure and P_c is the cuff pressure. Most methods that employ the occlusive arm cuff rely on this principle and differ in the means of detecting whether the artery is open or closed. Briefly, some approaches are the skin flush, palpatory, Korotkoff (auscultatory), oscillometric, and ultrasound methods [Drzewiecki et al., 1987]. Of these, the methods of Korotkoff and oscillometry are in most common use and will be reviewed here.

The idea of using lumen opening and closure as an indication of blood pressure has survived since its introduction by Marey. This simple concept is complicated by the fact that lumen closure may not necessarily occur at zero transmural pressure. Instead, transmural pressure must be negative by 5 to

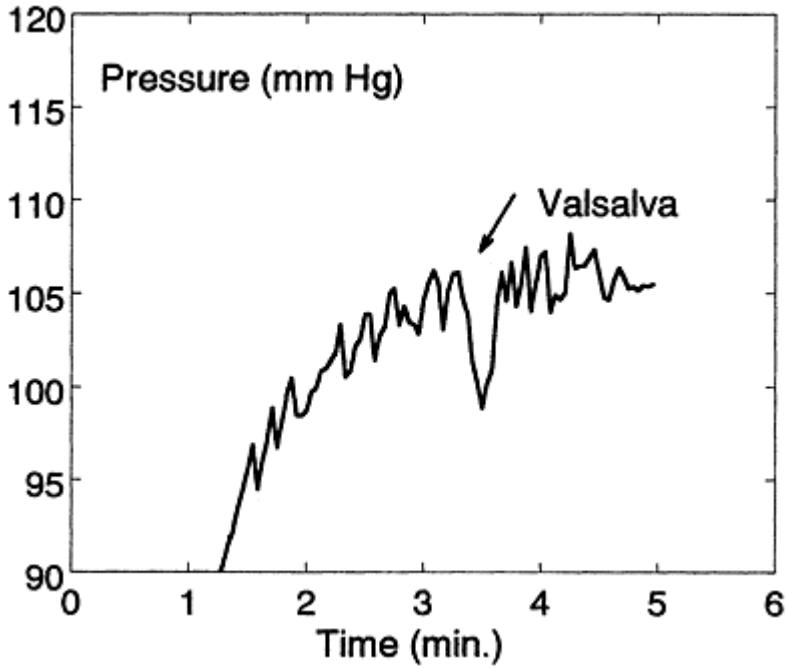


FIGURE 15.1 Application of arterial buckling to the continuous measurement of arterial pressure. The record shown tracks the subject's mean pressure in the brachial artery. The initial portion of the record illustrates the system as it locates the buckling point and the subject's blood pressure. The subject was directed to perform a brief Valsalva maneuver midway through the recording.

20 mmHg for complete closure. *In vitro* and *in vivo* human measurements have revealed that vessel buckling more closely approximates zero transmural pressure [Drzewiecki et al., 1997; Drzewiecki and Pilla, in press]. Buckling may be defined as the point at which the vessel switches from wall stretch to wall bending as a means of supporting its pressure load. The vessel is maximally compliant at this point and is approximately 25% open.

The validity of employing the buckling concept to blood pressure determination was tested. Sheth and Drzewiecki [1998] employed feedback control to regulate the pressure in a flexible diaphragm tonometer (see the section on *flexible diaphragm tonometer*). The buckling point was determined from the volume pulse. Thus, to detect vessel buckling, the derivative of the volume pulse with respect to mean pressure was computed. This was performed on a beat-to-beat basis. The feedback system was then used to adjust pressure such that this derivative was maximized, indicating the point of greatest instability. Thus, the underlying blood vessel was maintained in a constant state of buckling. According to the buckling theory, the transmural pressure should be nearly zero and tonometer pressure is equal to arterial pressure. A sample 2-min recording of noninvasive arterial pressure by this approach is shown (Fig. 15.1). The method was capable of tracking the subject's blood pressure in response to a Valsalva maneuver in this same record. This test demonstrates the feasibility of employing buckling to measure blood pressure. This example also illustrates that beat-to-beat pressure can be obtained by this method without the necessity to occlude blood flow. This approach should be useful for pressure variability studies.

Occlusive Cuff Mechanics

The occlusive arm cuff has evolved more out of convenience than engineering design. As such, its mechanical properties are relatively unknown. The current version of the cuff is designed to encircle the upper arm. It consists of a flat rubber bladder connected to air supply tubing. The bladder is covered externally by cloth material with Velcro fasteners at either end for easy placement and removal. While the cuff encircles the entire arm, the bladder extends over approximately half the circumference. The bladder is pressurized with air derived from a hand pump, release valve, and manometer connected to the air supply tubing.

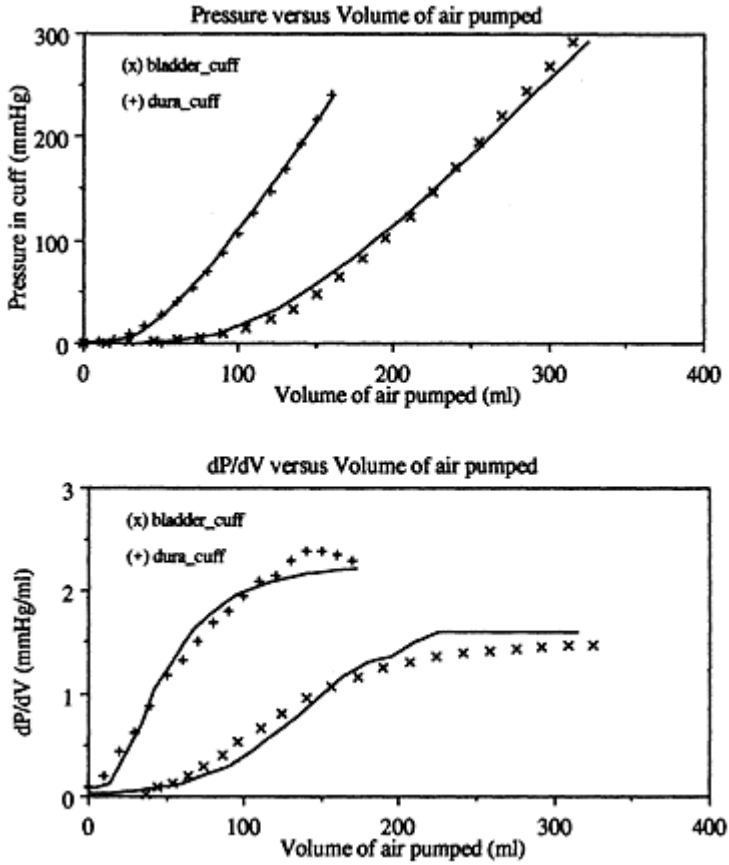


FIGURE 15.2 *Top.* Pressure-volume data obtained from two different occlusive arm cuffs. Inner surface of the cuff was fixed in this case to isolate cuff mechanics from that of the arm. *Bottom.* Derivative of the cuff pressure with respect to volume obtained from pressure-volume data of both cuffs. These curves indicate the pressure response of the cuff to volume change and are useful for plethysmography. Solid curves in both figures are the results of the occlusive cuff model.

The cuff should accurately transmit pressure down to the tissue surrounding the brachial artery. A mechanical analysis revealed that the length of the cuff is required to be a specific fraction of the arm's circumference for pressure equilibrium [Alexander et al., 1977]. A narrow cuff resulted in the greatest error in pressure transmission and, thus, the greatest error in blood pressure determination. Geddes and Whistler [1978] experimentally examined the effect of cuff size on blood pressure accuracy for the Korotkoff method. Their measurements confirmed that a cuff-width-to-arm circumference ratio of 0.4 should be maintained. Cuff manufacturers, therefore, supply a range of cuff sizes appropriate for pediatric use up to large adults.

Another aspect of the cuff is its pressure response due to either internal air volume change or that of the limb. In this sense, the cuff can be thought of as a plethysmograph. It was examined by considering its pressure-volume characteristics. In an experiment, the cuff properties were isolated from that of the arm by applying it to a rigid cylinder of similar diameter [Drzewiecki et al., 1993]. Pressure-volume data were then obtained by injecting a known volume of air and noting the pressure. This was performed for a standard bladder cuff (13-cm width) over a typical range of blood pressure.

The cuff pressure-volume results for pressures less than 130 mmHg are nonlinear (Fig. 15.2). For higher pressures, the data asymptotically approach a linear relationship. The cuff volume sensitivity, i.e., its derivative with respect to volume, increased with cuff pressure (Fig. 15.2). Above 130 mmHg, the cuff responded with nearly constant sensitivity.

A cuff mechanics theory was developed to explain the above cuff experiment [Drzewiecki et al., 1993]. Cuff mechanics was theorized to consist of two components. The first consists of the compressibility of air within the cuff. This was modeled using Boyle's gas law. The second component consists of elastic and geometric deformation of the cuff bladder. Cuff shape deformation proceeds at low pressures until the bladder reaches its final geometry, rendering a curved pressure-volume relationship. Then, elastic stretch of the rubber bladder takes over at high pressures, resulting in a nearly linear relationship. Solutions for this model are shown in comparison with the data in Fig. 15.2 for two cuffs; a standard cuff and the Critikon Dura-cuf. This model was useful in linearizing the cuff for use as a plethysmograph and for application to oscillometry (below).

Method of Korotkoff

The auscultatory method or method of Korotkoff was introduced by the Russian army physician N. Korotkoff [1905]. In his experiments, Korotkoff discovered that sound emitted distally from a partially occluded limb. He realized that this sound was indicative of arterial flow and that together with the occlusive cuff could be used to determine blood pressure. The method, as employed today, utilizes a stethoscope placed distal to an arm cuff over the brachial artery at the antecubital fossa. The cuff is inflated to about 30 mmHg above systolic pressure and then allowed to deflate at a rate of 2 to 3 mmHg/s. With falling cuff pressure, sounds begin and slowly change their characteristics. The initial "tapping" sounds are referred to as Phase I Korotkoff sound, and denote systolic pressure. The sounds increase in loudness during Phase II. The maximum intensity occurs in Phase III, where the tapping sound may be followed by a murmur due to turbulence.

Finally, Phase IV Korotkoff sound is identified as muffled sound, and Phase V is the complete disappearance of sound. Phase IV is generally taken to indicate diastolic arterial pressure. But, Phase V has also been suggested to be a more accurate indication of diastolic pressure. This matter is a continued source of controversy. The long history of the Korotkoff sound provides much experimental information. For example, the frequency spectrum of sound, spatial variation along the arm, filtering effects, and timing are reviewed [Drzewiecki et al., 1989].

It is a long-held misconception that the origin of the Korotkoff sound is flow turbulence. Turbulence is thought to be induced by the narrowing of the brachial artery under an occlusive cuff as it is forced to collapse. There are arguments against this idea. First, the Korotkoff sounds do not sound like turbulence, i.e., a murmur. Second, the Korotkoff sound can occur in low blood-flow situations, while turbulence cannot. And, last, Doppler ultrasound indicates that peak flow occurs following the time occurrence of Korotkoff sound. An alternative theory suggests that the sound is due to nonlinear distortion of the brachial pulse, such that sound is introduced to the original pulse. This is shown to arise from flow limitation under the cuff in addition to curvilinear pressure-area relationship of the brachial artery [Drzewiecki et al., 1989]. Strong support for this theory comes from its ability to predict many of the Korotkoff sound's observable features.

The accuracy of the Korotkoff method is well known. London and London [1967] find that the Korotkoff method underestimates systolic pressure by 5 to 20 mmHg and overestimates diastolic pressure by 12 to 20 mmHg. However, certain subject groups, such as hypertensives or the elderly, can compound these errors [Spence et al., 1978]. In addition, it has been shown that the arm blood flow can alter the Korotkoff sound intensity and, thus, the accuracy of blood pressure measurement [Rabbany et al., 1993]. Disappearance of Korotkoff sound occurs early in Phase III for some subjects and is referred to as the auscultatory gap. This causes an erroneous indication of elevated diastolic pressure. The auscultatory gap error can be avoided by simply allowing cuff pressure to continue to fall, where the true Phase IV sounds return. This is particularly critical for automatic instruments to take into account. In spite of these errors, the method of Korotkoff is considered a documented noninvasive blood pressure standard by which other noninvasive methods may be evaluated [White et al., 1993].

The Korotkoff method is applicable to other vessels besides the brachial artery of the arm. For example, the temporal artery has been employed [Shenoy et al., 1993]. In this case, a pressure capsule is applied over the artery on the head in place of an occlusive cuff, to provide external pressure. This approach has been shown to be accurate and is applicable to aerospace. Pilots' cerebral vascular pressure often falls in high-acceleration maneuvers, so that temporal artery pressure is a better indicator of this response.

Oscillometry

Oscillometric measurement of blood pressure predates the method of Korotkoff. The French physiologist Marey [1885] placed the arm within a compression chamber and observed that the chamber pressure fluctuated with the pulse. He also noted that the amplitude of pulsation varied with chamber pressure. Marey believed that the maximum pulsations or the onset of pulsations were associated with equality of blood pressure and chamber pressure. At that time, he was not certain what level of arterial pressure the

maximum pulsations corresponded with. Recently, it has been demonstrated theoretically that the variation in cuff pressure pulsation is primarily due to the brachial artery buckling mechanics [Drzewiecki et al., 1994].

Today, oscillometry is performed using a standard arm cuff together with an in-line pressure sensor. Due to the requirement of a sensor, oscillometry is generally not performed manually, but, rather, with an automatic instrument. The recorded cuff pressure is high-pass-filtered above 1 Hz to observe the pulsatile oscillations as the cuff slowly deflates (Fig. 15.3). It has been determined only recently that the maximum oscillations actually correspond with cuff pressure equal to mean arterial pressure (MAP) [Posey et al., 1969; Ramsey, 1979], confirming Marey's early idea. Systolic pressure is located at the point where the oscillations, O_s, are a fixed percentage of the maximum oscillations, O_m [Geddes et al., 1983]. In comparison with the intra-arterial pressure recordings, the systolic detection ratio is $O_s/O_m=0.55$. Similarly, the diastolic pressure can be found as a fixed percentage of the maximum oscillations, as $O_d/ O_m=0.85$.

Derivative Oscillometry

The use of the oscillometric detection ratios to find systolic and diastolic pressure is an empirical approach. That is, the ratios are statistically valid over the population of subjects that form the total

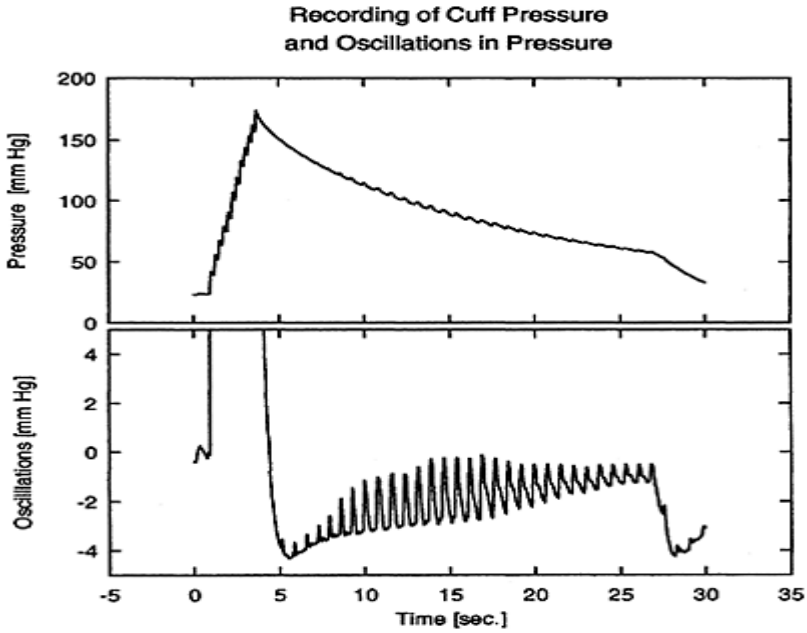


FIGURE 15.3 Sample recording of cuff pressure during oscillometric blood pressure measurement. Bottom

panel shows oscillations in cuff pressure obtained by high pass filtering above 1/2 Hz.

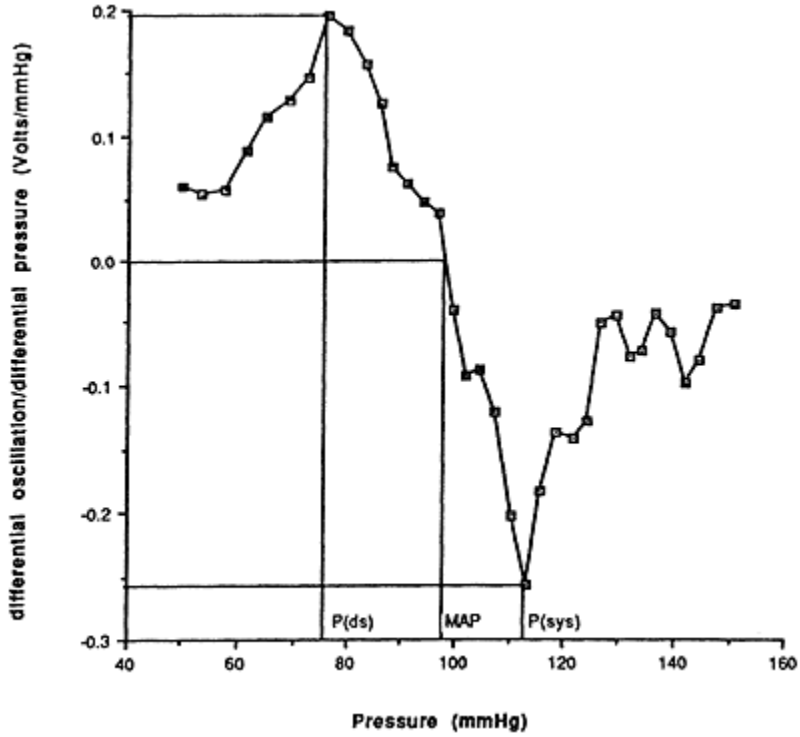


FIGURE 15.4 Method of derivative oscillometry. The derivative of cuff pressure oscillations data with respect to cuff pressure is shown from a single subject. The maximum and minimum values denote diastolic and systolic pressure, respectively. A zero derivative indicates MAP in this plot.

sample. This is a distinctly different approach than measuring blood pressure by detecting an event, such as the maximum in cuff pressure oscillations or the occurrence of Korotkoff sound. The event is constrained by the physics of arterial collapse under the cuff [Drzewiecki et al., 1994]. Therefore, it is more likely to be accurate under different conditions and, more importantly, for subjects outside of the sample population. In fact,

the oscillometric determination of MAP is more accurate than systolic and diastolic pressures.

Drzewiecki et al. [1994] employed a model of oscillometry to evaluate the derivative of the oscillation amplitude curve (Fig. 15.3) with respect to cuff pressure. When this derivative is plotted against cuff pressure, it was found that it reaches a maximum positive value. This occurred when cuff pressure equals diastolic pressure. Additionally, the minimum negative value was found to occur at systolic pressure. A measurement performed in our lab on a single subject illustrates the approach (Fig. 15.4). The specific advantage offered is that the empirically based systolic and diastolic ratios are not necessary [Link, 1987]. This method may be referred to as derivative oscillometry.

Derivative oscillometry was evaluated experimentally in our lab. The values of systolic and diastolic pressures obtained by derivative oscillometry were compared with those obtained by the method of Korotkoff. Thirty recordings were obtained on normal subjects (Fig. 15.5). The results indicated a high correlation of 0.93 between the two methods. Systolic mean error was determined to be 9% and diastolic mean error was −6%. Thus, derivative oscillometry was found to compare well with the method of Korotkoff in this preliminary evaluation. Before adopting derivative oscillometry, a more complete evaluation needs to be performed using a greater and more diverse subject population.

15.3 Pulse Dynamics Methods

15.3.1 R-Wave Time Interval Technique

One of the basic characteristics of pressure and flow under an occlusive cuff is that the pulse is apparently delayed with increasing cuff pressure. The R-wave of the EGG is often employed as a time

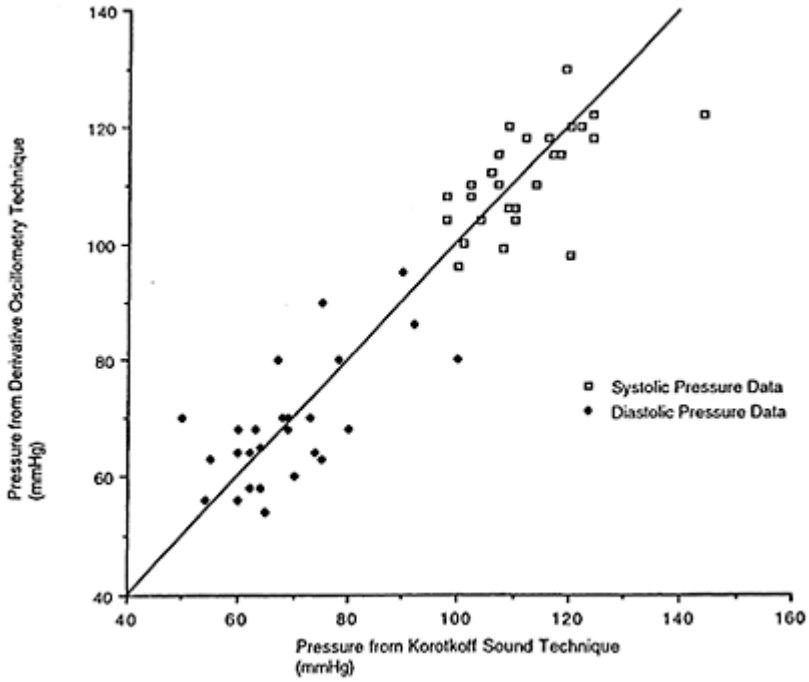


FIGURE 15.5 Experimental evaluation of derivative oscillometry using systolic and diastolic pressure from the method of Korotkoff as reference. The line indicates the result of linear regression to the data.

reference. Arzbaecher et al. [1973] measured the time delay of the Korotkoff sound relative to the R-wave. They suggested that the curve obtained by plotting this data represents the rising portion of the arterial pulse waveform.

A Korotkoff sound model [Drzewiecki et al., 1989] was employed to investigate the cuff delay effect. Time delay of the Korotkoff sound was computed relative to the proximal arterial pulse. Since the arterial pulse waveform was known in this calculation, the pressure-RK interval curve can be compared directly. The resemblance to a rising arterial pulse was apparent but some deviation was noted, particularly in the early portion of the RK interval curve. The model indicates that the pulse occurs earlier and with higher derivative than the true pulse waveform. In particular, the increased derivative that occurs at the foot of the pulse may mislead any study of wave propagation.

Recently, Sharir et al. [1993] performed comparisons of Doppler flow pulse delay and direct arterial pressure recordings. Their results confirm a consistent elevation in pulse compared with intra-arterial recording in the early portions of the wave. The average deviation was 10 mmHg in value. Hence, if accuracy is not important, this method can

provide a reasonable estimate of blood pressure and its change. The commercial pulse watch has been a recent application of this approach.

Continuous Vascular Unloading

Penaz [1973] reasoned that if the cuff pressure could be continuously adjusted to equal the arterial pressure, the vessels would be in a constant state of vascular unloading. He employed mechanical feedback to continuously adjust the pressure in a finger chamber to apply this concept. The vascular volume was measured using photoplethysmography. When feedback was applied such that the vascular volume was held constant and at maximum pulsatile levels, the chamber pressure waveform was assumed to be equal to the arterial pressure.

Recent applications of the Penaz method have been developed by Wesseling et al. [1978] and Yamakoshi et al. [1980]. One instrument is commercially available as the FINAPRES (Ohmeda, Finapres, Englewood, CO). These instruments have been evaluated in comparison with intra-arterial pressure recordings [Omboni et al., 1993]. Good waveform agreement has been obtained in comparison with intra-arterial measurements from the radial artery. But, it should be clear that the Penaz method employs finger pressure, which is a peripheral vascular location. This recording site is prone to pulse wave reflection effects and is therefore sensitive to the vascular flow resistance. It is anticipated that the technique would be affected by skin temperature, vasoactive drugs, and anesthetics. Moreover, mean pressure differences between the finger pulse and central aortic pressure should be expected.

Pulse Sensing

Pulse sensors attempt to measure the arterial pulse waveform from either the arterial wall deflection or force at the surface of the skin above a palpable vessel. Typically, these sensors are not directly calibrated in terms of pressure, but ideally respond proportionately to pressure. As such, they are primarily useful for dynamic information. While there are many designs available for this type of sensor, they generally fall into two categories. The first category is that of the volume sensor (Fig. 15.6 [left]). This type of sensor relies on the adjacent tissues surrounding the vessel as a nondeflecting frame of reference as, for example, a photoplethysmograph. Skin deflections directly above the vessel are then measured relative to a reference frame to represent the arterial pressure. Several different transduction methods may be employed such as capacitive, resistive, inductive, optical, etc. Ideally, this type of sensor minimally restricts the motion of the skin so that contact force is zero. The drawback to pulse volume sensing is that the method responds to volume distention and *indirectly* to pressure. The nonlinear and viscoelastic nature of the vascular wall result in complex waveform alterations for this type of sensor that are difficult to correct in practice.

The second category is that of the pressure pulse sensor (Fig. 15.6 [right]). This type of sensor measures stress due to arterial pressure transmitted though the skin above the pulse artery. The pressure pulse sensor requires that surface deflections are zero, as opposed to the volume sensor. Thus, the contact forces are proportionate to arterial pressure at the skin surface.

The differences in pulse waveforms that are provided by the above pulse recording techniques are clear in comparison with intra-arterial recordings [Van der Hoeven and Beneken, 1970]. In all cases, the pressure pulse method was found to provide superior waveform accuracy, free of the effects of vascular nonlinear viscoelasticity. Alternatively, the stiffness of the sensor can best characterize its pulse accuracy. High stiffness relative to the artery and surrounding tissue is required to best approximate the pressure pulse method.

Arterial pulse recording is performed while the subject is stationary and refrains from moving the pulse location. But, it has become of interest to acquire ambulatory records. Without restraint, pulse recording is quite difficult due to motion artifact. For example, hand or finger motion can appear in recordings of the radial artery. A change in sensor positioning or acceleration can result in other types of artifact. The artifacts are often comparable in magnitude and frequency to the pulse, rendering simple filtering methods useless. Recently though, artifact cancellation techniques that employ sensor arrays have been applied with good success [Ciaccio et al., 1989].

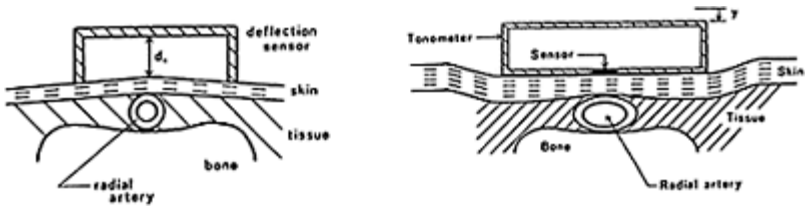


FIGURE 15.6 *Left.* Illustration of volume pulse method. *Right.* Illustration of pressure pulse method and arterial tonometry.

Arterial Tonometry

Pulse-sensing methods do not provide calibrated pressure. Arterial tonometry [Pressman and Newgard, 1963] is a pressure pulse method that can noninvasively record calibrated pressure in superficial arteries with sufficient bony support, such as the radial artery.

A tonometer is applied by first centering a contact stress sensor over the vessel. This is accomplished by repositioning the device until the largest pulse is detected. An array of sensors [Weaver et al., 1978] has been used to accomplish this electronically. Then, the tonometer is depressed toward the vessel. This leads to applanation of the vessel wall (Fig. 15.6). If the vessel is not flattened sufficiently, the tonometer measures forces due to arterial wall tension and bending of the vessel. As depression is continued, the arterial wall is applanated further, but not so much as to occlude blood flow. At this intermediate position, wall tension becomes parallel to the tonometer sensing surface. Arterial pressure is then the remaining stress perpendicular to the surface and is measured by the sensor. This is termed the contact stress due to pressure. Ideally, the sensor should not measure skin shear (frictional) stresses. The contact stress is equal in magnitude to the arterial pressure when these conditions are achieved. The details of arterial tonometer calibration

and design were analyzed by Drzewiecki et al. [1983, 1987]. In summary, tonometry requires that the contact stress sensor be flat, stiffer than the tissues, and small relative to the vessel diameter. Proper calibration can be attained either by monitoring the contact stress distribution (using a sensor array) or the maximum in measured pulse amplitude.

Recent research in tonometry has focused on miniaturization of semiconductor pressure sensor arrays [Weaver et al., 1978]. Alternatively, fiber optics have been employed by Drzewiecki [1985] and Moubarak et al. [1989], allowing extreme size reduction of the contact stress sensor. Commercial technology has been available (Jentow, Colin Electronics, Japan) and has been evaluated against intra-arterial records. Results indicate an average error of −5.6 mmHg for systolic pressure and −2.4 mmHg for diastole. Excellent pulse waveform quality is afforded by tonometry [Sato et al., 1993], making it a superior method for noninvasive pulse dynamics applications.

Flexible Diaphragm Tonometry

As an alternative to the high-resolution tonometers under development, a new low-resolution technology is introduced here. The basic advantage becomes one of cost, ease of positioning, and patient comfort, which is critical for long-term applications. In addition, while most tonometers have employed only the radial artery of the wrist for measurement, this technology is suitable for other superficial vessels and conforms to skin surface irregularities.

The flexible tonometer design applies the fact that tissue is incompressible in the short term [Bansal et al., 1994]. The concept is shown in Fig. 15.7 using three tonometer volume compartments. These compartments are not physically separated and fluid can move between them. They are coupled to the skin and artery by means of the flexible diaphragm. When the arterial pressure exceeds that in the tonometer, the volume of the artery expands into Vb. Note also, that Va and Vc must increase to take up this expansion since water is incompressible. To restore the tonometer to a flat surface, the total volume of the tonometer is increased (Fig. 15.7). In response, the tonometer pressure increases and the artery flattens. At this point, the volume in each compartment is equal, the diaphragm is flat, and the tonometer pressure is equal to arterial pressure. Thus, by maintaining the equilibrium of the relative volume compartments, applanation tonometry can be accomplished with a flexible diaphragm rather than a rigid one. In practice, instrumentation continuously adjusts the relative compartment volumes as arterial pressure changes.

A flexible diaphragm tonometer was machined from plexiglass (Fig. 15.8). A rectangular channel was designed to contain saline. The front of the channel was sealed with a polyurethane sheet of 0.004 in. thickness. Two stainless steel electrodes were placed at each end of the channel. These were used to inject a current along the channel length. Near the center of the channel, four measuring electrodes were placed at equal spacing. Each pair of electrodes defined a volume compartment and

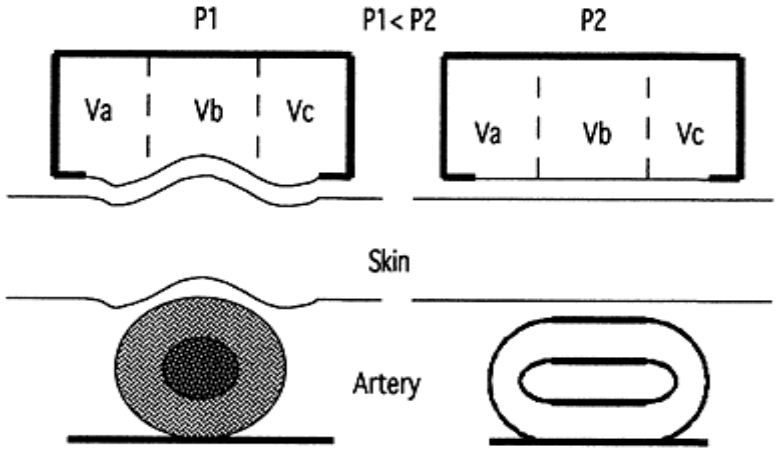


FIGURE 15.7 Concept of flexible diaphragm tonometry. P1 illustrates inappropriate level of pressure to provide arterial applanation tonometry. P2 indicates an increase in chamber volume so that the relative compartment volumes, V, are equal. In practice, relative compartment volumes are maintained constant via feedback control and applanation is continuous. Compartment pressure equals arterial pressure in P2.

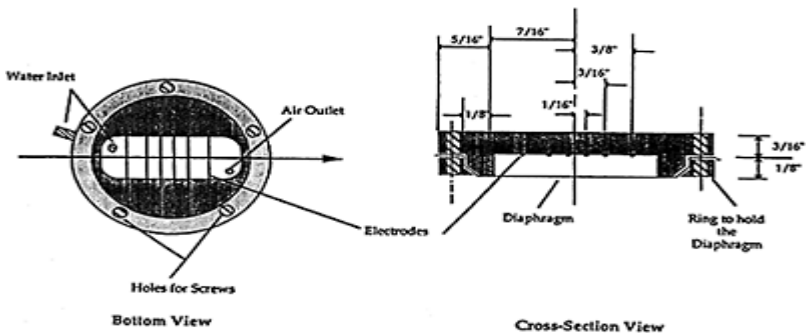


FIGURE 15.8 Design of a flexible diaphragm tonometer in accordance with the concept shown in Fig. 15.7.

Compartment volumes are obtained by means of impedance plethysmography. Electrode positions define the compartment boundaries.

the voltage across each pair was calibrated in terms of volume using impedance plethysmography. External to the tonometer, a saline-filled catheter was used to connect the channel to an electromechanical volume pump. The dynamics of the system were designed to possess appropriate frequency response.

Several beats of data were plotted from the volume pulse recordings of the flexible tonometer (Fig. 15.9) for two adjacent compartments. The pulse volume is shown as a volume deviation from the mean compartment volume given a constant average tonometer pressure. The waveform shown illustrates the ability to provide calibrated noninvasive arterial volume information.

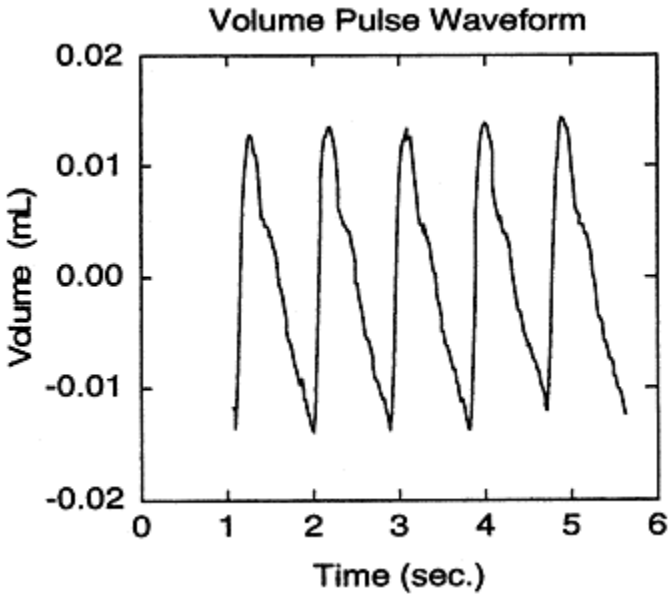


FIGURE 15.9 Sample of radial arterial volume recording from tonometer of Fig. 15.8. Chamber pressure was fixed at 25 mmHg. Calibrated volume change is shown relative to the mean compartment volume.

15.4 Noninvasive Arterial Mechanics

The flexible diaphragm tonometer and the occlusive cuff are capable of measuring volume in addition to pressure. Combining noninvasive measurements can extend the information provided by a single instrument, beyond that of blood pressure alone. Furthermore, functional information about the vasculature requires the simultaneous determination of pressure as well as geometry. In this section, the standard occlusive arm cuff will be employed as a plethysmograph to illustrate this concept.

The transmural pressure vs. lumen area (P-A) graph has been a convenient approach to analyze vessel function [Drzewiecki et al., 1997]. This curve can be applied to hemodynamics and portrays vessel function in a single graph. Unfortunately, its use has been limited to invasive studies or studies of isolated vessels where pressure can be experimentally controlled. The occlusive arm cuff offers a noninvasive solution to this problem by allowing the transmural pressure to be varied by altering the external pressure applied to a blood vessel. This approach has been used to noninvasively measure the transmural pressure vs. compliance relationship of the brachial artery [Mazhbich et al., 1983].

Drzewiecki and Pilla [1998] furthered the idea of noninvasively varying the transmural pressure by developing a cuff-based instrument that measures the P-A relationship. To perform the measurement, an occlusive arm cuff was pressurized by means of a diaphragm air pump. Cuff pressure was measured using a pressure sensor that provides pressure as an electronic signal. The pressure signal was input to a computer system via an analog-to-digital converter for analysis. Two valves were used to control the flow of air. Initially, the valves were operated so that the pump inflated the cuff to well above the subject's systolic pressure. At that point, cuff air was recirculated through the pump. Under this condition, the mean cuff pressure remained constant, but the stroke volume of the pump resulted in cuff pressure oscillations at the frequency of the pump. Since the pump volume is a known quantity, it was used as a volume calibration source. The pump volume was divided by the cuff pressure oscillations due to the pump to yield the cuff compliance. With cuff compliance known, the cuff pressure arterial pulse was converted to a volume pulse by multiplying by cuff compliance. The pump was designed to operate at approximately 40 Hz, well above the heart rate frequency components. This allowed the arterial pulse and pump pulse to be easily separated from the cuff pressure recording by using low pass and high pass digital filters with a cutoff frequency of 25 Hz.

The above procedure was repeated at every level of cuff pressure. The cuff pressure was released in steps of 5 to 10 mmHg until it was zero by briefly opening the release valve. The subject's pulse volume was divided by their pressure pulse (systolic minus diastolic pressures) to obtain the arterial compliance at every value of cuff pressure. The compliance per unit length was obtained by further dividing by the effective cuff length. The systolic, diastolic, and mean arterial pressures were evaluated by oscillometry and the Korotkoff method. Arterial compliance was then plotted for every value of transmural pressure (Fig. 15.10). The compliance was then numerically integrated to obtain the corresponding brachial artery P-A curve. Since the vessel was collapsed for large negative transmural pressures, the initial constant of integration was chosen to be zero lumen area.

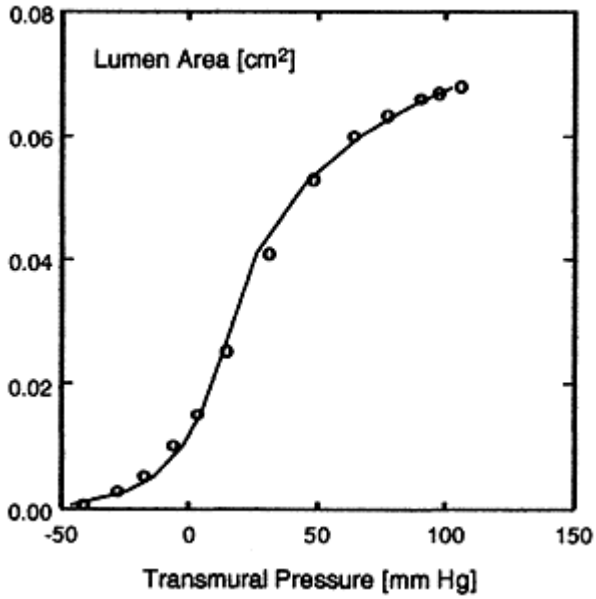


FIGURE 15.10 Arterial compliance obtained noninvasively by employing the occlusive arm cuff as a plethysmograph. Points are measured data. The solid curve is the model result obtained from Eq. (15.1).

The arterial compliance curve possesses a maximum near zero transmural pressure. This point corresponds with the onset of vessel buckling. On either side of the buckling point, the vessel supports its pressure load differently. On the negative pressure side, the vessel partially collapses and undergoes wall bending. The compliance rapidly approaches zero during collapse. This occurs as the opposite walls of the vessel begin to contact each other and close the lumen. On the positive pressure side, the vessel supports its pressure load by wall stretch. The compliance slowly decreases with increasing pressure because of nonlinear wall elasticity. That is, the wall becomes stiffer with increasing stretch. The s-shaped lumen area curve is a consequence of these two different mechanisms. Preliminary studies of several subjects revealed that the shape of the noninvasive P-A curve is consistent for all subjects studied [Whitt and Drzewiecki, 1997; Drzewiecki and Pilla, 1998].

A mathematical model was developed that incorporates the fundamental physical and material properties that contribute to the collapsible P-A relationship [Drzewiecki et al., 1997; Drzewiecki and Pilla, 1998]. The basic form of the model is summarized by the following equation for transmural pressure:

$$P = -E\left(\left(\lambda^{-1}\right)^{n} - 1\right) + P_b + a\left(e^{b(\lambda-1)} - 1\right), \qquad (15.1)$$

where a and b are arterial elastance constants for distension, E is vessel elastance during collapse, and n is a constant that determines the rate of change of pressure with respect to change in area during collapse. The quantity λ, was defined as the extension ratio and is evaluated from the lumen area divided by the lumen area at the buckling point, A/A_b. The first hyperbolic term is the pressure due to collapse and wall bending. The second term is the buckling pressure and is found when $A=A_b$. The third exponential term represents the pressure due to wall stretch. Some overlap in the contribution of each term may occur near the buckling pressure. Depending on the vessel type and material, Eq. (15.1) can be improved by limiting the extension ratio and its inverse to unity.

Equation (15.1) was employed to analyze the brachial artery P-A data from each subject. The constants A_b and P_b were measured directly from the data as the point on the P-A curve that corresponds with maximum compliance or buckling. Their values were inserted into Eq. (15.1) for each subject, leaving the remaining constants to be found by nonlinear least squares regression (Marquardt-Levenberg algorithm). The model was evaluated for the subject shown in Fig. 15.10 and corresponds with the solid line curve. Similar results were obtained for all subjects studied (N=10), with the mean error of estimate less than 3 mmHg. The above study suggests that no other physical properties need to be added to model vascular collapse. Thus, it can be considered a valid model for further studies of vascular properties and blood pressure determination.

While the other terms of the model were found to vary from subject to subject, the buckling pressure was discovered to be relatively constant (10 mmHg±11). Hence, buckling may be the important vascular feature that permits noninvasive blood pressure measurement to be feasible and may be the common thread that links all noninvasive methods. This idea was first examined in our theoretical examination of oscillometry above [Drzewiecki et al., 1994]. The successful use of buckling itself as a method to find arterial pressure was also described here [Sheth and Drzewiecki, 1998]. The phenomenon of buckling is a general effect that occurs independent of the technique used to find arterial pressure. It will also be independent of the quantitative differences in the P-A relationship of each specific subject.

15.5 Summary

Future research is open to noninvasive studies of how cardiovascular disease can alter blood vessel mechanics and the accuracy of blood pressure determination. The use of methods, such as occlusive cuff plethysmography together with blood pressure measurement and vascular mechanical modeling presented here, offers a means for noninvasive detection of the early stages of cardiovascular disease. Additionally, pulse dynamics and pressure variation offered by noninvasive pulse recording can provide new information about cardiovascular control.

Marey originally proposed that vessel closure is the important event that permits noninvasive blood pressure measurement. From the work presented here, this early

concept is refocused to the instability of arterial buckling, or the *process* of closure, as the basic mechanism that enables noninvasive blood pressure determination.

Acknowledgments

The author wishes to express thanks to graduate students Vineet Bansal, Edward Ciaccio, Cindy Jacobs, James Pilla, Deepa Sheth, and Michael Whitt for their assistance with the experimental results presented here. The oscillometry studies were supported, in part, by Critikon, Inc., Tampa, FL, and arterial tonometry was supported, in part, by IVAC corp., San Diego, CA.

References

Alexander H, Cohen M, and Steinfeld L. 1977. Criteria in the choice of an occluding cuff for the indirect measurement of blood pressure. *Med. Biol Eng. Comput.* 15:2–10.

Arzbaecher RC and Novotney RL. 1973. Noninvasive measurement of the arterial pressure contour in man. *Bibl Cardiol.* 31:63–69.

Bansal V, Drzewiecki G, and Butterfield R. 1994. Design of a flexible diaphragm tonometer. *Thirteenth S. Biomed. Eng. Conf.* Washington, D.C. pp. 148–151.

Ciaccio EJ, Drzewiecki GM, and Karam E. 1989. Algorithm for reduction of mechanical noise in arterial pulse recording with tonometry. *Proc. 15th Northeast Bioeng. Conf.* p. 161–162, Boston.

Drzewiecki GM. 1985. The Origin of the Korotkoff Sound and Arterial Tonometry. Ph.D. dissertation. University of Pennsylvania, Philadelphia.

Drzewiecki G, Field S, Moubarak I, and Li J K-J. 1997. Vascular growth and collapsible pressure-area relationship. *Amer. J. Physiol* 273:H2030-H2043.

Drzewiecki G, Hood R, and Apple H. 1994. Theory of the oscillometric maximum and the systolic and diastolic detection ratios. *Ann. Biomed. Eng.* 22:88–96.

Drzewiecki GM, Karam E, Bansal V, Hood R, and Apple H. 1993. Mechanics of the occlusive arm cuff and its application as a volume sensor. *IEEE Trans. Biomed. Eng.* BME-40:704–708.

Drzewiecki GM, Melbin J, and Noordergraaf A. 1983. Arterial tonometry: Review and analysis. *J. Biomech.* 16:141–152.

Drzewiecki GM, Melbin J, and Noordergraaf A. 1987. Noninvasive blood pressure recording and the genesis of Korotkoff sound. In *Handbook of Bioengineering*, S Chien, R Skalak., eds. pp. 8.1–8.36. New York, McGraw-Hill.

Drzewiecki GM, Melbin J, and Noordergraaf A. 1989. The Korotkoff sound. Ann. *Biomed. Eng.* 17:325–359.

Drzewiecki G and Pilla J. 1998. Noninvasive measurement of the human brachial artery pressure—area relation in collapse and hypertension. *Ann. Biomed. Eng.* Nov–Dec; 26(6)965–74.

Graettinger WF, Lipson JL, Cheung DG, and Weber MA. 1988. Validation of portable noninvasive blood pressure monitoring devices: Comparisons with intra-arterial and sphygmomanometer measurements. *Am. Heart J.* 116:1155–1169.

Geddes LA, Voelz M, Combs C, and Reiner D. 1983. Characterization of the oscillometric method for measuring indirect blood pressure. *Ann. Biomed. Eng.* 10:271–280.

Geddes LA and Whistler SJ. 1978. The error in indirect blood pressure measurement with incorrect size of cuff. *Am. Heart J.* 96:4–8.

Kelly R, Daley J, Avolio A, and O'Rourke M. 1989. Arterial dilation and reduced wave reflection—Benefit of Dilevalol in Hypertension. *Hypertension.* 14:14–21.

Korotkoff N. 1905. On the subject of methods of determining blood pressure. *Bull Imperial Mil. Med. Acad.* (St. Petersburg) 11:365–367.
Li J K-J. 1986. Time domain resolution of forward and reflected waves in the aorta. *IEEE Trans. Biomed. Eng.* BME-33:783–785.
Link WT. 1987. Techniques for obtaining information associated with an individual's blood pressure including specifically a stat mode technique. US Patent #4,664,126.
London SB and London RE. 1967. Comparison of indirect blood pressure measurements (Korotkoff) with simultaneous direct brachial artery pressure distal to cuff. *Adv. Intern. Med.* 13:127–142.
Marey EJ. 1885. La *Methode Graphique* dans les Sciences Experimentales et Principalement en Physiologic et en Medicine. Masson, Paris.
Maurer A and Noordergraaf A. 1976. Korotkoff sound filtering for automated three-phase measurement of blood pressure. *Am. Heart J.* 91:584–591.
Mazhbich BJ. 1983. Noninvasive determination of elastic properties and diameter of human limb arteries. *Pflugers Arch.* 396:254–259.
Moubarak IF, Drzewiecki GM, and Kedem J. 1989. Semi-invasive fiber—Optic tonometer. *Proc. 15th Northeast Bioeng. Conf.* p. 167–168.
Noordergraaf A. 1978. *Circulatory System Dynamics.* Academic Press, New York.
Omboni S, Parati G, Frattol A, Mutti E, Di Rienzo M, Castiglioni P, and Mancia G. 1993. Spectral and sequence analysis of finger blood pressure variability: Comparison with analysis of intra-arterial recordings. *Hypertension.* 22:26–33.
Pickering T, James G, Boddie C, Harshfield G, Blank S, and Laragh J. 1988. How common is white coat hypertension? *JAMA.* 259:225–228.
Posey JA, Geddes LA, Williams H, and Moore AG. 1969. The meaning of the point of maximum oscillations in cuff pressure in the indirect measurement of blood pressure. Part 1. *Cardiovasc. Res. Cent. Bull.* 8:15–25.
Pressman GL and Newgard PM. 1963. A transducer for the continuous external measurement of arterial blood pressure. *IEEE Trans. Biomed. Eng.* BME-10:73–81.
Rabbany SY, Drzewiecki GM, and Noordergraaf A. 1993. Peripheral vascular effects on auscultatory blood pressure measurement. *J. Clin. Monitoring.* 9:9–17.
Ramsey III M. 1979. Noninvasive blood pressure determination of mean arterial pressure. *Med. Biol Engng.* Comp., 17:11–18.
Penaz J. 1973. Photoelectric measurement of blood pressure, volume, and flow in the finger. *Dig. 10th Intl. Conf. Med. Eng.* Dresden, Germany, p. 104.
Sato T, Nishinaga M, Kawamoto A, Ozawa T, and Takatsuji H. 1993. Accuracy of a continuous blood pressure monitor based on arterial tonometry. *Hypertension.* 21:866–874.
Sharir T, Marmor A, Ting C-T, Chen J-W, Liu C-P, Chang M-S, Yin FCP, and Kass DA. 1993. Validation of a method for noninvasive measurement of central arterial pressure. *Hypertension.* 21:74–82.
Shenoy D, von Maltzahn WW, and Buckley JC. 1993. Noninvasive blood pressure measurement on the temporal artery using the auscultatory method. *Ann. Biomed. Eng.* 21:351–360.
Sheth D and Drzewiecki G. 1998. Using vessel buckling for continuous determination of arterial blood pressure. *Ann. Biomed. Eng.* 26:S-70.
Spence JD, Sibbald W J, and Cape RD. 1978. Pseuodohypertension in the elderly. *Clin. Sci. Mol Med.* 55:399s-402s.
Strus J. 1555. Sphygmicae artis jam mille ducentos annos peritae et desideratae, Libri V a Josephi Struthio Posnanience, medico recens conscripti, Basel. (As transcribed by A. Noordergraaf, Univ. of Pennsylvania, Philadelphia, PA).
Van der Hoeven GMA and Beneken JEW. 1970. A reliable transducer for the recording of the arterial pulse wave. *Prog. Rep. 2. Inst. Med. Physics.* TNO, Utrecht.

Wesseling KH, de Wit B, Snoeck B, Weber JAP, Hindman BW, Nijland R, and Van der Hoeven GMA. 1978. An implementation of the Penaz method for measuring arterial blood pressure in the finger and the first results of an evaluation. *Prog. Rep. 6. Inst. Med. Phys.* TNO, Utrecht.

Weaver CS, Eckerle JS, Newgard PM, Warnke CT, Angell JB, Terry SC and Robinson J. 1978. A study of noninvasive blood pressure measurement technique. In Noninvasive Cardiovascular Measurements. *Soc. Photo-Optical Instr. Eng.* 167:89.

Westerhof N, Bosman F, DeVries CJ, and Noordergraaf A. 1969. Analog studies of the human systemic arterial tree. *J. Biomech.* 2:121–143.

White WW, Berson AS, Robbins C, Jamieson MJ, Prisant M, Roccella E, and Sheps SG. 1993. National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers. *Hypertension.* 21:504–509.

Whitt M and Drzewiecki G. 1997. Repeatability of brachial artery area measurement. *Ann. Biomed. Eng.* 25:8–12.

Yamakoshi K, Shimazu H, and Togawa T. 1980. Indirect measurement of instantaneous arterial blood pressure in the human finger by the vascular unloading technique. *IEEE Trans. Biomed. Eng.* BME-27:150.

16
Cardiac Output Measurement

Leslie A.Geddes
Purdue University

Cardiac output is the amount of blood pumped by the right or left ventricle per unit of time. It is expressed in liters per minute (L/min) and normalized by division by body surface area in square meters (m^2). The resulting quantity is called the cardiac index. Cardiac output is sometimes normalized to body weight, being expressed as mL/min per kilogram. A typical resting value for a wide variety of mammals is 70 mL/min per kg.

With exercise, cardiac output increases. In well-trained athletes, cardiac output can increase five-fold with maximum exercise. During exercise, heart rate increases, venous return increases, and the ejection fraction increases. Parenthetically, physically fit subjects have a low resting heart rate, and the time for the heart rate to return to the resting value after exercise is less than that for subjects who are not physically fit.

There are many direct and indirect (noninvasive) methods of measuring cardiac output. Of equal importance to the number that represents cardiac output is the left-ventricular ejection fraction (stroke volume divided by diastolic volume), which indicates the ability of the left ventricle to pump blood.

16.1 Indicator-Dilution Method

The principle underlying the indicator-dilution method is based on the upstream injection of a detectable indicator and on measuring the downstream concentration-time curve, which is called a *dilution curve*. The essential requirement is that the indicator mixes with all the blood flowing through the central mixing pool. Although the dilution curves in the outlet branches may be slightly different in shape, they all have the same area.

Figure 16.1*a* illustrates the injection of *m* g of indicator into an idealized flowing stream having the same velocity across the diameter of the tube. Figure 16.1*b* shows the dilution curve recorded downstream. Because of the flow-velocity profile, the cylinder of indicator and fluid becomes teardrop in shape, as shown in Fig. 16.1*c*. The resulting dilution curve has a rapid rise and an exponential fall, as shown in Fig. 16.1*d*. However, the area of the dilution curve is the same as that shown in Fig. 16.1*a*. Derivation of the flow equation is shown in Fig. 16.1, and the flow is simply the amount of indicator (*m* gm) divided by the area of the dilution curve (gm/mL×s), which provides the flow in mL/s.

0-8493-1813-0/03/$0.00+$1.50

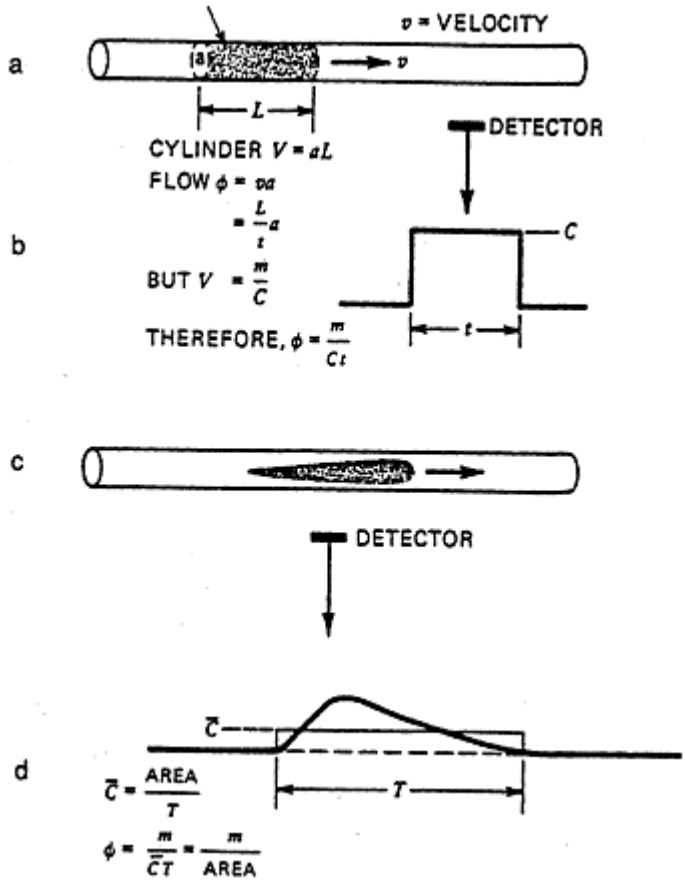


FIGURE 16.1 Genesis of the indicator-dilution curve.

Indicators

Before describing the various indicator-dilution methods, it is useful to recognize that there are two types of indicator, diffusible and nondiffusible. A diffusible indicator will leak out of the capillaries. A nondiffusible indicator is retained in the vascular system for a time that depends on the type of indicator. Whether cardiac output is overestimated with a diffusible indicator depends on the location of the injection and measuring sites. Table 16.1 lists many of the indictors that have been used for measuring cardiac output and the types of detectors used to obtain the dilution curve. It is obvious that the indicator selected must be detectable and not alter the flow being measured. Importantly, the indicator must be nontoxic and sterile.

When a diffusible indicator is injected into the right heart, the dilution curve can be detected in the pulmonary artery, and there is no loss of indicator because there is no capillary bed between these sites; therefore the cardiac output value will be accurate.

TABLE 16.1 Indicators

Material	Detector	Retention Data
Evans blue (T1824)	Photoelectric 640 mu.	50% loss in 5 days
Indocyanine green	Photoelectric 800 mu.	50% loss in 10 minutes
Coomassie blue	Photoelectric 585–600 mu.	50% loss in 15–20 minutes
Saline (5%)	Conductivity cell	Diffusible*
Albumin 1131	Radioactive	50% loss in 8 days
Na^{24}, K^{42}, D_2O, DHO	Radioactive	Diffusible*
Hot-cold solutions	Thermodetector	Diffusible*

* It is estimated that there is about 15% loss of diffusible indicators during the first pass through the lungs.

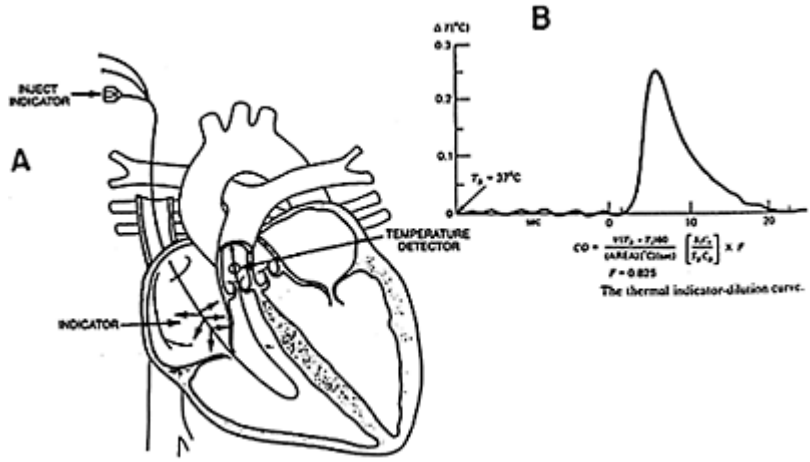


FIGURE 16.2 The thermodilution method *(a)* and a typical dilution curve *(b)*.

Thermal Dilution MethodÅ

Chilled 5% dextrose in water (D5W) or 0.9% NaCl can be used as indicators. The dilution curve represents a transient reduction in pulmonary artery blood temperature following injection of the indicator into the right atrium. Figure 16.2 illustrates the method and a typical thermodilution curve. Note that the indicator is really negative

calories. The thermodilution method is based on heat exchange measured in calories, and the flow equation contains terms for the specific heat *(C)* and the specific gravity *(S)* of the indicator *(i)* and blood *(b)*. The expression employed when a #7F thermistor-tipped catheter is used and chilled D5W is injected into the right atrium is as follows:

$$\mathrm{CO} = \left[\frac{V\left(T_b - T_i\right)60}{A}\right]\left[\frac{S_i C_i}{S_b C_b}\right]F,$$

where V=Volume of indicator injected in mL

T_b=Temperature (average of pulmonary artery blood in (°C)

T_i=Temperature of the indicator (°C)

60=Multiplier required to convert mL/s into mL/min

A=Area under the dilution curve in (seconds×°C)

S=Specific gravity of indicator *(i)* and blood *(b)*

C=Specific heat of indicator *(i)* and blood *(b)*

$\left(\frac{S_i C_i}{S_b C_b} = 1.08 \text{ for 5\% dextrose and blood of 40\% packed-cell volume}\right)$

F=Empiric factor employed to correct for heat transfer through the injection catheter (for a #7F catheter, F=0.825 [2]).

Entering these factors into the expression gives

$$\mathrm{CO} = \frac{V\left(T_b - T_i\right)53.46}{A},$$

where CO=cardiac output in mL/min

53.46=60×1.08×0.825.

To illustrate how a thermodilution curve is processed, cardiac output is calculated below using the dilution curve shown in Fig. 16.2.

V=5 ml of 5% dextrose in water

T_b=37°C

T_i=0°C

A=1.59°C s

$$\mathrm{CO} = \frac{5\left(37 - 0\right)53.46}{1.59} = 6220 \text{ mL/min.}$$

CO= 6220mL/min.

Although the thermodilution method is *the standard in clinical medicine*, it has a few disadvantages. Because of the heat loss through the catheter wall, several series 5-mL injections of indicator are needed to obtain a consistent value for cardiac output. If cardiac output is low, i.e., the dilution curve is very broad, it is difficult to obtain an accurate value for cardiac output. There are respiratory-induced variations in PA blood temperature that confound the dilution curve when it is of low amplitude. Although room-temperature D5W can be used, chilled D5W provides a better dilution curve and a

more reliable cardiac output value. Furthermore, it should be obvious that if the temperature of the indicator is the same as that of blood, there will be no dilution curve.

Indicator Recirculation

An ideal dilution curve, shown in Fig. 16.2, consists of a steep rise and an exponential decrease in indicator concentration. Algorithms that measure the dilution-curve area have no difficulty with such a curve. However, when cardiac output is low, the dilution curve is typically low in amplitude and very broad. Often the descending limb of the curve is obscured by recirculation of the indicator or by low-amplitude artifacts. Figure 16.3a is a dilution curve in which the descending limb is obscured by recirculation of the indicator. Obviously it is difficult to determine the practical end of the curve, which is often specified as the time when the indicator concentration has fallen to a chosen percentage (e.g., 1%) of the maximum amplitude (C_{max}). Because the descending limb represents a good approximation of a decaying exponential curve (e^{-kt}), fitting the descending limb to an exponential allows reconstruction of the curve without a recirculation error, thereby providing a means for identifying the end for what is called the *first pass of the indicator*.

In Fig. 16.3*b*, the amplitude of the descending limb of the curve in Fig. 16.3*a* has been plotted on semilogarithmic paper, and the exponential part represents a straight line. When recirculation appears, the data points deviate from the straight line and therefore can be ignored, and the linear part (representing the exponential) can be extrapolated to the desired percentage of the maximum concentration, say 1% of C_{max}. The data points representing the extrapolated part were replotted on Fig. 16.3*a* to reveal the dilution curve undistorted by recirculation.

Commercially available indicator-dilution instruments employ digitization of the dilution curve. Often the data beyond about 30% of C_{max} are ignored, and the exponential is computed on digitally extrapolated data.

16.2 Fick Method

The Fick method *employs oxygen as the indicator* and the increase in oxygen content of venous blood as it passes through the lungs, along with the respiratory oxygen uptake, as the quantities that are needed to determine cardiac output, $CO=O_2$ uptake/$A—VO_2$ difference). Oxygen uptake (mL/min) is measured at the airway, usually with an oxygen-filled spirometer containing a CO_2 absorber. The $A—VO_2$ difference is determined from the oxygen content (mL/100 mL blood) from any arterial sample and the oxygen

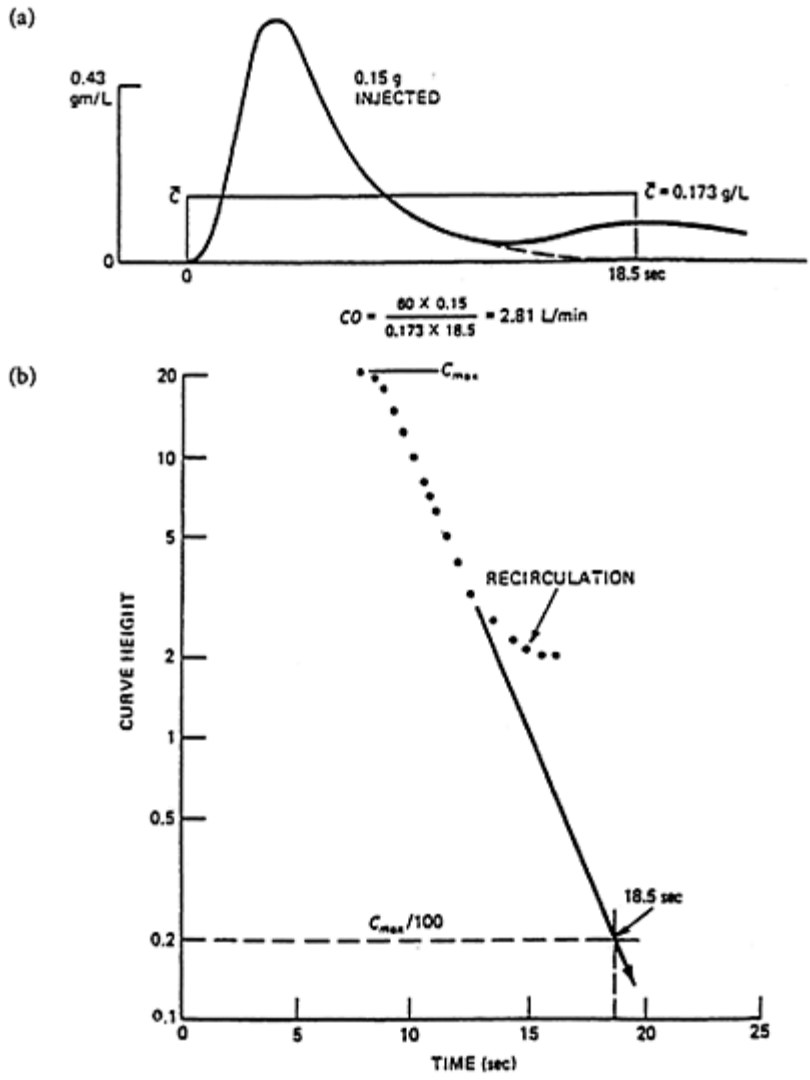


FIGURE 16.3 Dilution curve obscured by recirculation *(a)* and a semilogarithmic plot of the descending limb *(b)*.

content (mL/100 mL) of pulmonary arterial blood. The oxygen content of blood used to be difficult to measure. However, the new blood-gas analyzers that measure, pH, pO_2, pCO_2, hematocrit, and hemoglobin provide a value for O_2 content by computation using the oxygen-dissociation curve.

There is a slight technicality involved in determining the oxygen uptake because oxygen is consumed at body temperature but measured at room temperature in the spirometer. Consequently, the volume of O_2 consumed per minute displayed by the spirometer must be multiplied by a factor, *F*. Therefore the Pick equation is

$$CO = \frac{O_2 \text{ uptake}/\min(F)}{A - VO_2 \text{ difference}}.$$

Figure 16.4 is a spirogram showing a tidal volume riding on a sloping baseline that represents the resting expirating level (REL). The slope identifies the oxygen uptake at room temperature. In this subject, the uncorrected oxygen consumption was 400 mL/min at 26°C in the spirometer. With a barometric pressure of 750 mmHg, the conversion factor *F* to correct this volume to body temperature (37°C) and saturated with water vapor is

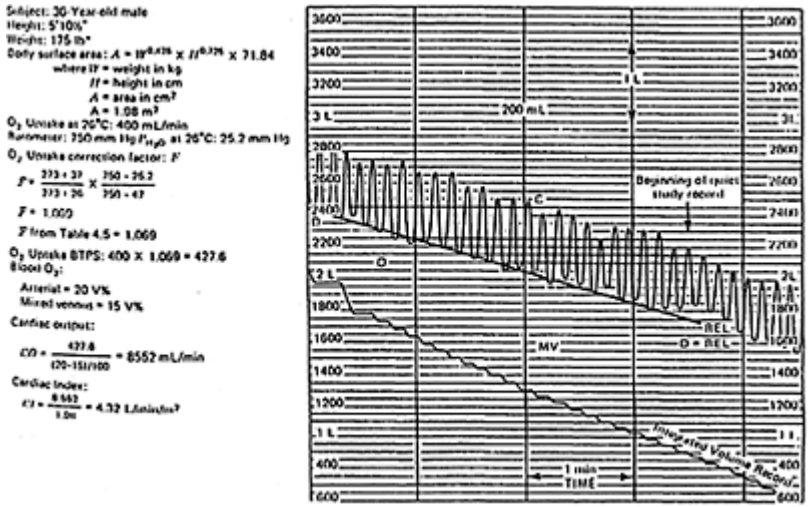


FIGURE 16.4 Measurement of oxygen uptake with a spirometer *(right)* and the method used to correct the measured volume *(left)*.

$$F = \frac{273+37}{273+T_s} \times \frac{P_b - PH_2O}{P_b - 47},$$

where T_i is the spirometer temperature, P_b is the barometric pressure, and PH_2O at T_s is obtained from the water-vapor table (Table 16.2).

A sample calculation for the correction factor *F* is given in Fig. 16.4, which reveals a value for *F* of 1.069. However, it is easier to use Table 16.3 to obtain the correction

factor. For example, for a spirometer temperature of 26°C and a barometric pressure of 750 mmHg, F=1.0691.

Note that the correction factor F in this case is only 6.9%. The error encountered by not including it may be less than the experimental error in making all other measurements.

The example selected shows that the $A—VO_2$ difference is 20–15 mL/100 mL blood and that the corrected O_2 uptake is 400×1.069; therefore the cardiac output is:

$$CO = \frac{400 \times 1.069}{(20-15)/100} = 8552 \text{ mL/min}.$$

The Fick method does not require the addition of a fluid to the circulation and may have value in such a circumstance. However, its use requires stable conditions because an average oxygen uptake takes many minutes to obtain.

16.3 Ejection Fraction

The ejection fraction (EF) is one of the most convenient indicators of the ability of the left (or right) ventricle to pump the blood that is presented to it. Let v be the stroke volume (SV) and V be the end-diastolic volume (EDV); the ejection fraction is v/V or SV/EDV.

TABLE 16.2 Vapor Pressure of Water

Temp.°C	0.0	0.2	0.4	0.6	0.8
15	12.788	12.953	13.121	13.290	13.461
16	13.634	13.809	13.987	14.166	14.347
17	14.530	14.715	14.903	15.092	15.284
18	15.477	15.673	15.871	16.071	16.272
19	16.477	16.685	16.894	17.105	17.319
20	17.535	17.753	17.974	18.197	18.422
21	18.650	18.880	19.113	19.349	19.587
22	19.827	20.070	20.316	20.565	20.815
23	21.068	21.324	21.583	21.845	22.110
24	22.377	22.648	22.922	23.198	23.476
25	23.756	24.039	24.326	24.617	24.912
26	25.209	25.509	25.812	26.117	26.426
27	26.739	27.055	27.374	27.696	28.021
28	28.349	28.680	29.015	29.354	29.697

29	30.043	30.392	30.745	31.102	31.461
30	31.825	32.191	32.561	32.934	33.312
31	33.695	34.082	34.471	34.864	35.261
32	35.663	36.068	36.477	36.891	37.308
33	37.729	38.155	38.584	39.018	39.457
34	39.898	40.344	40.796	41.251	41.710
35	42.175	42.644	43.117	43.595	44.078
36	44.563	45.054	45.549	46.050	46.556
37	47.067	47.582	48.102	48.627	49.157
38	49.692	50.231	50.774	51.323	51.879
39	42.442	53.009	53.580	54.156	54.737
40	55.324	55.910	56.510	57.110	57.720
41	58.340	58.960	59.580	60.220	60.860

Measurement of ventricular diastolic and systolic volumes can be achieved radiographically, ultrasonically, and by the use of an indicator that is injected into the left ventricle where the indicator concentration is measured in the aorta on a beat-by-beat basis.

Indicator-Dilution Method for Ejection Fraction

Holt [1] described the method of injecting an indicator into the left ventricular during diastole and measuring the stepwise decrease in aortic concentration with successive beats (Fig. 16.5). From this concentration-time record, end-diastolic volume, stroke volume, and ejection fraction can be calculated. No assumption need be made about the geometric shape of the ventricle. The following describes the theory of this fundamental method.

Let V be the end-diastolic ventricular volume. Inject m gm of indicator into this volume during diastole. The concentration (C_1) of indicator in the aorta for the first beat is m/V. By knowing the amount of indicator (m) injected and the calibration for the aortic detector, C_1 is established, and ventricular end-diastolic volume $V=m/C_1$.

After the first beat, the ventricle fills, and the amount of indicator left in the left ventricle is $m—mv/V$. The aortic concentration (C_2) for the second beat is therefore $m—mV/V=m(l—v/V)$. Therefore the aortic concentration (C_2) for the second beat is

$$C_2 = \frac{m}{V}\left[1 - \frac{v}{V}\right].$$

By continuing the process, it is easily shown that the aortic concentration (C_n) for the nth beat is

TABLE 16.3 Correction Factor *F* for Standardization of Collected Volume

°C/P_B	640	650	660	670	680	690	700	710	720	730	740	750	760	770	780
15	1.1388	1.1377	1.1367	1.1358	1.1348	1.1339	1.1330	1.1322	1.1314	1.1306	1.1298	1.1290	1.1283	1.1276	1.1269
16	1.1333	1.1323	1.1313	1.1304	1.1295	1.1286	1.1277	1.1269	1.1260	1.1253	1.1245	1.1238	1.1231	1.1224	1.1217
17	1.1277	1.1268	1.1266	1.1249	1.1240	1.1232	1.1224	1.1216	1.1208	1.1200	1.1193	1.1186	1.1179	1.1172	1.1165
18	1.1222	1.1212	1.1203	1.1194	1.1186	1.1178	1.1170	1.1162	1.1154	1.1147	1.1140	1.1133	1.1126	1.1120	1.1113
19	1.1165	1.1156	1.1147	1.1139	1.1131	1.1123	1.1115	1.1107	1.1100	1.1093	1.1086	1.1080	1.1073	1.1067	1.1061
20	1.1108	1.1099	1.1091	1.1083	1.1075	1.1067	1.1060	1.1052	1.1045	1.1039	1.1032	1.1026	1.1019	1.1094	1.1008
21	1.1056	1.1042	1.1034	1.1027	1.1019	1.1011	1.1004	1.0997	1.0990	1.0984	1.0978	1.0971	1.0965	1.0960	1.0954
22	1.0992	1.0984	1.0976	1.0969	1.0962	1.0964	1.0948	1.0941	1.0935	1.0929	1.0923	1.0917	1.0911	1.0905	1.0900
23	1.0932	1.0925	1.0918	1.0911	1.0904	1.0897	1.0891	1.0884	1.0878	1.0872	1.0867	1.0861	1.0856	1.0850	1.0845
24	1.0873	1.0866	1.0859	1.0852	1.0846	1.0839	1.0833	1.0827	1.0822	1.0816	1.0810	1.0805	1.0800	1.0795	1.0790
25	1.0812	1.0806	1.0799	1.0793	1.0787	1.0781	1.0775	1.0769	1.0764	1.0758	1.0753	1.0748	1.0744	1.0739	1.0734
26	1.0751	1.0710	1.0738	1.0732	1.0727	1.0721	1.0716	1.0710	1.0705	1.0700	1.0696	1.0691	1.0686	1.0682	1.0678
27	1.0688	1.0682	1.0677	1.0671	1.0666	1.0661	1.0656	1.0651	1.0640	1.0641	1.0637	1.0633	1.0629	1.0624	1.0621
28	1.0625	1.0619	1.0614	1.0609	1.0604	1.0599	1.0595	1.0591	1.0586	1.0582	1.0578	1.0574	1.0570	1.0566	1.0563
29	1.0560	1.0555	1.0550	1.0546	1.0548	1.0537	1.0533	1.0529	1.0525	1.0521	1.0518	1.0514	1.0519	1.0507	1.0504
30	1.0494	1.0496	1.0486	1.0482	1.0478	1.0474	1.0470	1.0467	1.0463	1.0460	1.0450	1.0453	1.0450	1.0447	1.0444

Source: From Kovach JC, Paulos P, Arabadjis C. 1955. *J Thorac Surg* **29**:552. *Vs*=FV_C, where V_s is the standardized condition and V_c is the collected condition:

$$V_s = \frac{1+37/273}{1+t°C/273} \times \frac{P_B - PH_2O}{P_B - 47} V_c = FV_c .$$

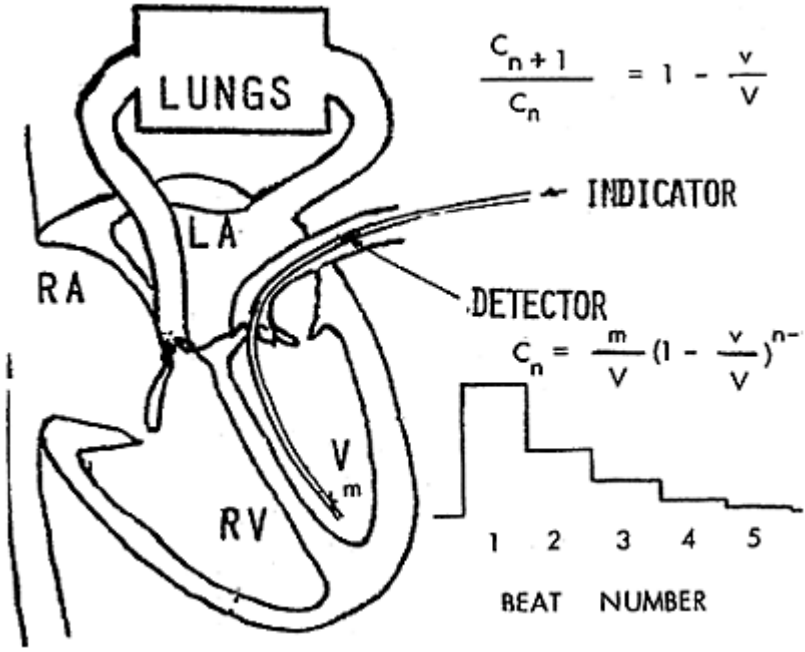


FIGURE 16.5 The saline method of measuring ejection fraction, involving injection of *m* gm of NaCl into the left ventricle and detecting the aortic concentration *(C)* on a beat-by-beat basis.

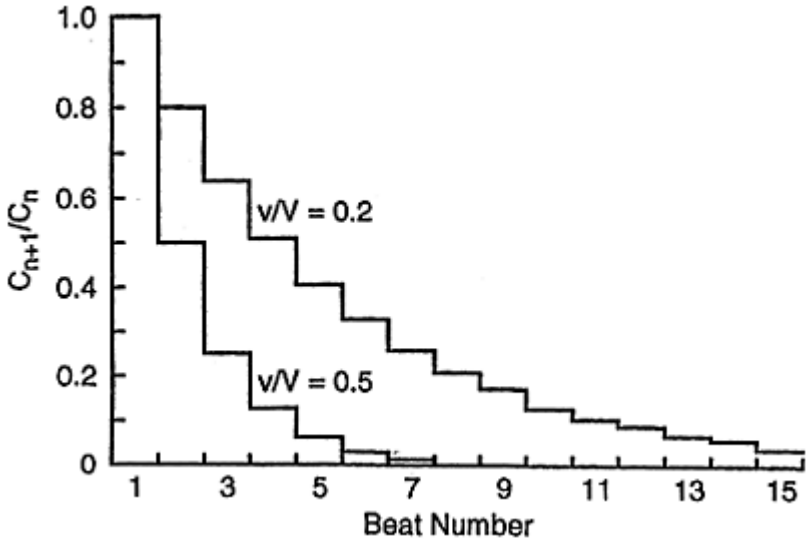


FIGURE 16.6 Stepwise decrease in indicator concentration (C) versus beat number for ejection fraction *(v/V)* of 0.5 and 0.2.

$$C_n = \frac{m}{V}\left[1 - \frac{v}{V}\right]^{n-1}.$$

Figure 16.6 illustrates the stepwise decrease in aortic concentration for ejection fractions *(v/V)* of 0.2 and 0.5, i.e., 20% and 50%.

It is possible to determine the ejection fraction from the concentration ratio for two successive beats. For example,

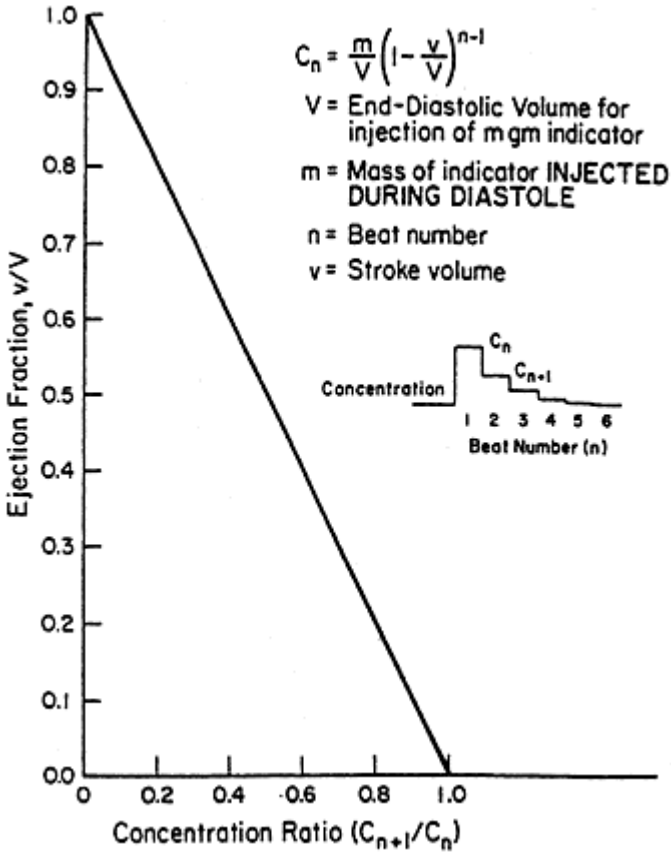


FIGURE 16.7 Ejection fraction (v/V) versus the ratio of concentrations for successive beats *(C_{n+1}/C_n)*.

$$C_n = \frac{m}{V}\left[1-\frac{v}{V}\right]^{n-1}$$

$$C_n + 1 = \frac{m}{V}\left[1-\frac{v}{V}\right]^{n}$$

$$\frac{C_{n+1}}{C_n} = 1-\frac{v}{V}.$$

from which

$$\frac{v}{V} = 1 - \frac{C_{n+1}}{C_n},$$

where v/V is the ejection fraction and C_{n+1}/C_n is the concentration ratio for two successive beats, e.g., C_2/C_1 *or* C_3/C_2. Figure 16.7 illustrates the relationship between the ejection fraction v/V and the ratio of C_{n+1}/C_n. Observe that the detector need not be calibrated as long as there is a linear relationship between detector output and indicator concentration in the operating range.

References

1. Holt JP. 1956. Estimation of the residual volume of the ventricle of the dog heart by two indicator-dilution techniques. *Circ Res* 4:181.
2. Weissel RD, Berger RL, Hechtman HB. 1975. Measurement of cardiac output by thermodilution. *N Engl J Med* 292:682.

17
Bioelectric Impedance Measurements

Robert Patterson
University of Minnesota

Bioelectric tissue impedance measurements to determine or infer biological information have a long history dating back to before the turn of the century. The start of modern clinical applications of bioelectric impedance (BEI) measurements can be attributed in large part to the reports by Nyboer [1970]. BEI measurements are commonly used in **apnea** monitoring, especially for infants, and in the detection of venous **thrombus**. Many papers report the use of the BEI technique for peripheral blood flow, cardiac stroke volume, and body composition. Commercial equipment is available for these latter three applications, although the reliability, validity, and accuracy of these applications have been questioned and, therefore, have not received widespread acceptance in the medical community.

BEI measurements can be classified into two types. The first and most common application is in the study of the small pulsatile impedance changes associated with heart and respiratory action. The goal of this application is to give quantitative and qualitative information on the volume changes (**plethysmography**) in the lung, heart, peripheral arteries, and veins. The second application involves the determination of body characteristics such as total body fluid volume, inter- and extracell volume, percent body fat, and cell and tissue viability. In this application, the total impedance is used and in some cases measured as a function of frequency, which is referred to as **impedance spectroscopy**.

17.1 Measurement Methods

Most single-frequency BEI measurements are in the range of 50 to 100 kHz (at these frequencies no significant electrical shock hazard exists) using currents from 0.5 to 4 mA RMS. Currents at these levels are usually necessary to obtain a good signal-to-noise ratio when recording the small pulsatile changes that are in the range of .1 to 1% of the total impedance. The use of higher frequencies creates instrumentation design problems due to stray capacity.

BEI measurements in the 50 to 100 kHz range have typical skin impedance values 2 to 10 times the value of the underlying body tissue of interest depending on electrode area. In order to obtain BEI values that can be used to give quantitative biological information, the skin impedance contribution must be eliminated. This is accomplished by using the

0-8493-1813-0/03/$0.00+$ 1.50

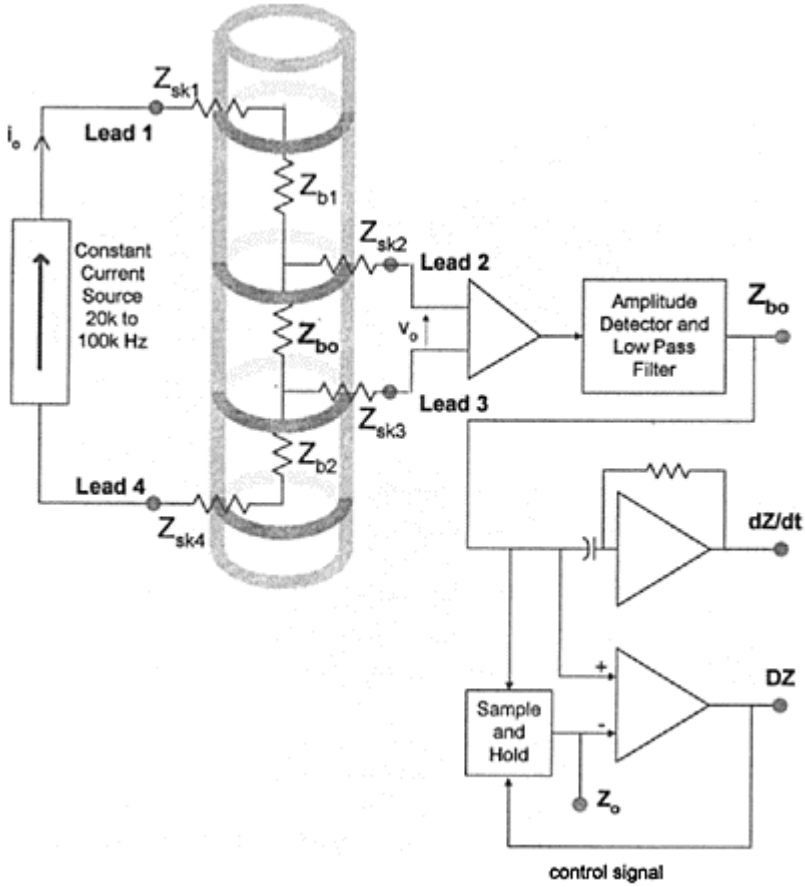


FIGURE 17.1 The four-electrode impedance measurement technique and the associated instrumentation.

four electrode impedance measurement method shown in Fig. 17.1, along with other signal processing blocks used in typical impedance plethysmographs.

Z_{bo} is the internal section of tissue we wish to measure. If we used two electrodes to make the measurement, we would include two skin impedances (i.e., Z_{sk1} and Z_{sk4}) and two internal tissue impedances (i.e., Z_{b1} and Z_{b2}), which would make it impossible to estimate an accurate value for Z_{bo}.

A constant current source supplies current, I_o, to the outside two electrodes 1 and 4. This current flows through the skin and body tissue independent of tissue and skin impedance values. The voltage V_o is measured across Z_{bo} with a voltage amplifier using electrodes 2 and 3. Assuming the output impedance of the current source is >> $Z_{sk1}+Z_{b1}+Z_{bo}+Z_{b2+}Z_{sk4}$ and the input impedance of the voltage amplifier is>>$Z_{sk2}+Z_{bo}+Z_{sk3}$, then

$$Z_{bo} = Z_o + \Delta Z, \quad Z_o = V_o / I_o, \quad \text{and} \quad \Delta Z = \Delta V_o / I_o, \tag{17.1}$$

where Z_o is the non-time-varying portion of the impedance and ΔZ is the impedance change typically-associated with the pulsation of blood in the region of measurement.

The output from the voltage pick-up amplifier (Fig. 17.1) is connected to the amplitude detector and low-pass filter that removes the high-frequency carrier signal, which results in an output voltage proportional to Z_{bo}. Z_{bo} has a large steady part that is proportional to the magnitude of the tissue impedance (Zo) and a small (.1 to 1%) part, ΔZ, which represents the change due to respiratory or cardiac activity. In order to obtain a signal representing ΔZ, Z_o must be removed from Z_{bo} and the signal amplified. This can be accomplished by capacity coupling or by subtracting a constant that represents Z_o. The latter is usually done because many applications require near dc response. The output of the ΔZ amplifier will be a waveform oscillating around zero volts. The output from the ΔZ amplifier controls the sample and hold circuit. When the ΔZ output exceeds a given value, usually plus or minus a few tenths of an ohm,

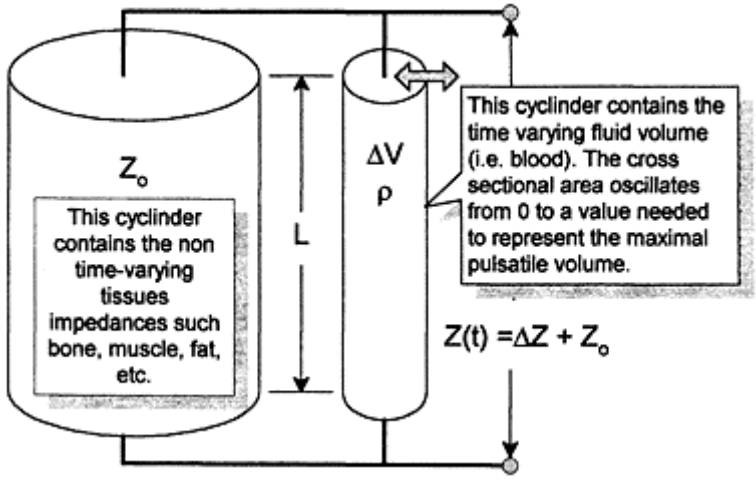


FIGURE 17.2 Parallel column model.

the sample and hold circuit updates its value of Z_o. The output from the sample and hold circuit is subtracted from Z_{bo} by the ΔZ amplifier. The derivative of Z_{bo} is frequently obtained in instruments intended for cardiac use.

17.2 Modeling and Formula Development

To relate the ΔZ obtained on the thorax or peripheral limbs to the pulsatile blood volume change, the parallel column model, first described by Nyboer [197o], is frequently used (Fig. 17.2). The model consists of a conducting volume with impedance Z_o in parallel

with a time-varying column with resistivity ρ, length L, and a time-varying cross-sectional area which oscillates from 0 to a finite value. At the time in the cardiac cycle when the pulsatile volume is at a minimum, all of the conducting tissues and fluids are represented by the volume labeled Z_o. This volume can be a heterogeneous mixture of all of the non-time-varying tissues such as fat, bone, muscle, etc. in the region under measurement. The only information needed about this volume is its impedance Z_o and that it is electrically in parallel with the small time-varying column. During the cardiac cycle, the volume change in the right column starts with a zero cross-sectional area and increases in area until its volume equals the blood volume change. If the impedance of this volume is much greater than Z_o, then the following relation holds:

$$\Delta V = \rho\left(L^2/Z_o^2\right)\Delta Z\,, \tag{17.2}$$

where ΔV=the pulsatile volume change with resistivity p

ρ=the resistivity of the pulsatile volume in Ω-cm (typically the resistivity of blood)

L=the length of the cylinder

Z_o=the impedance measured when the pulsatile volume is at a minimum

ΔZ=the magnitude of the pulsatile impedance change.

The resistivity of blood, p in Ω-cm, is a function of hematocrit (H) expressed as a percentage and can be calculated as ρ=67.919 exp(0.0247 H) [Mohapatra et al., 1977]. The typical value used for blood is 150 Ω-cm.

17.3 Respiration Monitoring and Apnea Detection

If the BEI is measured across the thorax, a variation of approximately 1 to 2 ohms per liter of lung volume change is observed, which increases with inspiration. The most common position of the electrodes for respiratory measurements is on each side of the thorax along the midaxillary line. The largest signal is

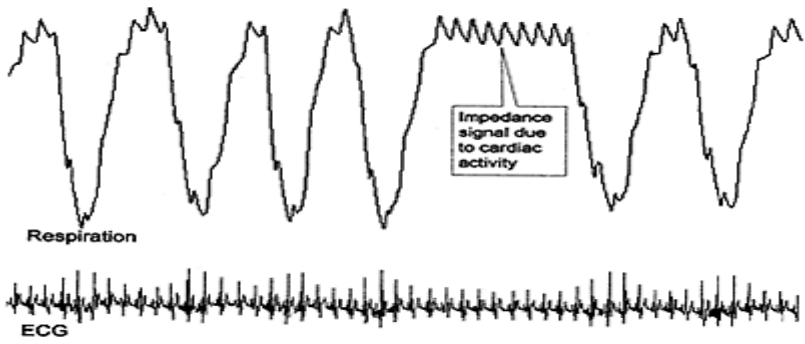


FIGURE 17.3 Example of BEI respiration signal and EGG.

generally obtained at the level of the xiphsternal joint, although a more linear signal is obtained higher up near the axilla [Geddes and Baker, 1989].

The problems encountered with the quantitative use of BEI for respiration volume are movement artifacts and the change in the response, depending on whether diaphragmatic or intercostal muscles are used. For most applications, the most serious problem is body movement and positional changes artifacts that can cause impedance changes significantly larger than the change caused by respiration.

The determination of apnea or whether respiration has stopped [Neuman, 1988] in infants is one of the most widely used applications of BEI. For convenience and due to the lack of space on the thorax of infants, only two electrodes are used. These are placed at the midthoracic level along the midaxillary line and are also used to obtain the EGG. No effort is usually made to quantitate the volume change. Filtering is used to reduce movement artifacts and automatic gain controls and adaptive threshold detection is used in the breath detection circuits. Due to movement artifacts, the normal breath detection rate in infants is not highly reliable. When respiration stops, body movement ceases, which eliminates the movement artifacts and then apnea can be detected. Ventation detection problems can occur if the airway is obstructed and the infant makes inspiratory movement efforts or cardiac-induced impedance changes are interpreted as a respiratory signal. Figure 17.3 shows a typical impedance measurement during an apneic period.

17.4 Peripheral Blood Flow

BEI measurements are made on limbs to determine arterial blood flow into the limb or for the detection of venous thrombosis. In both applications, an occluding cuff is inflated above venous pressure to prevent outflow for a short period of time.

Figure 17.4 shows the typical electrode arrangement on the leg and the position of the occluding cuff. The cuff is rapidly inflated to 40 to 50 mmHg, which prevents venous outflow without significantly changing arterial inflow. The arterial inflow causes an increase in the volume of the limb. The slope of the initial impedance change as determined by the first three or four beats is used to measure the arterial flow rate. Equation (17.2) is used to calculate the volume change from the impedance change. The flow (the slope of the line in Fig. 17.4) is determined by dividing the volume change by the time over which the impedance change was measured. The volume change that occurs after the impedance change reaches a plateau is a measure of the **compliance** of the venous system.

After the volume change of the leg has stabilized, the cuff is quickly deflated, which results in an exponential decrease in volume. If a thrombosis exists in the veins, the time constant of the outflow lengthens. The initial slope, the time constant, and percentage change at 3 sec after cuff deflation have been used to quantitate the measurement. The percentage change at 3 sec has been reported to show the

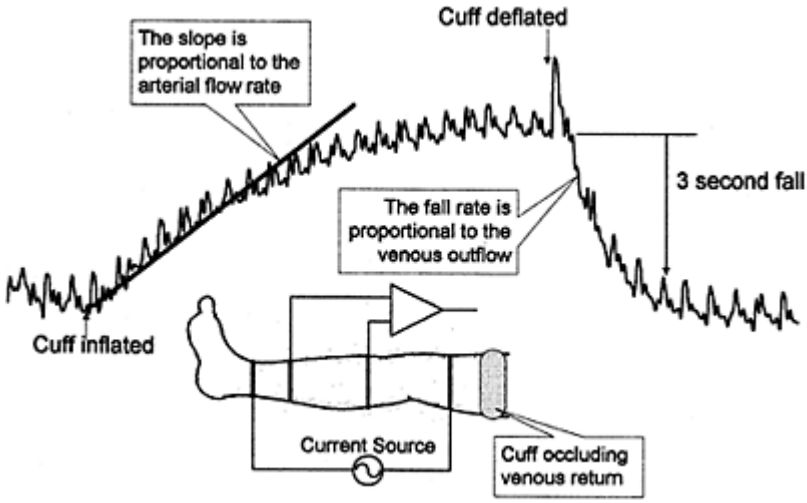


FIGURE 17.4 The measurement of arterial inflow and venous outflow.

best agreement with venograms. The determination of deep venous thrombus is frequently made by combining the maximal volume change with the outflow rate. The agreement with a venogram is 94% for the detection of deep venous thrombus proximal to the knee [Anderson, 1988].

17.5 Cardiac Measurements

The measurements of chest impedance changes due to cardiac activity have been reported starting in 1930s. One of the most popular techniques, first reported by Patterson et al. [1964], for quantitative measurements uses band electrodes around the ends of the thorax as shown in Fig. 17.5. Each heartbeat causes a pulsatile

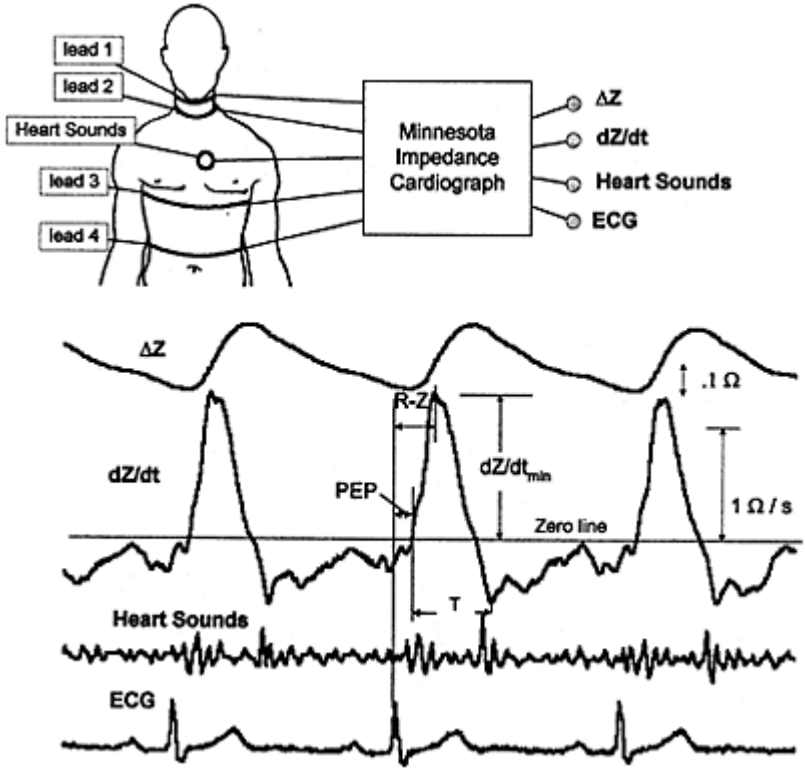


FIGURE 17.5 Impedance cardiographic waveforms.

decrease in impedance of 0.1 to 0.2 ohms (decreasing ΔZ and negative dZ/dt are shown in an upward direction). The empirical formula for stroke volume based on this method follows from Eq. (17.2):

$$\Delta V = \rho\left(L^2/Z_o^2\right) T \, dZ_{min}/dt \,, \tag{17.3}$$

where ΔV=cardiac stroke volume (mL)

ρ=resistivity of blood (Ω-cm)

L=distance between the inner band electrodes (cm)

Z_o=base impedance (Ω)

dZ_{min}/dt=the magnitude of the largest negative derivative of the impedance change occurring during systole (Ω/s)

T=systolic ejection time (seconds).

Many studies have been conducted comparing the impedance technique with other standard methods of measuring stroke volume and cardiac output. In general, the correlation coefficient in subjects without valve problems or heart failure and with a stable circulatory system is 0.8 to 0.9. In patients with a failing circulatory system or

valvular or other cardiovascular problems, the correlation coefficient may be less than 0.6 [Patterson, 1989].

Experimental physiological studies and computer modeling show that multiple sources contribute to the impedance signal. The anatomical regions that make significant contributions are the aorta, lungs, and atria. Recent studies have reported that the blood resistivity change with flow, pulsatile changes in the neck region, and the movement of the heart significantly contribute to the signal [Patterson et al., 1991, Wang and Patterson, 1995]. It appears that a number of different sources of the thoracic impedance change combine in a fortuitous manner to allow for a reasonable correlation between BEI measured stroke volume and other accepted techniques. However, in patients with cardiac problems where the contributions of the different sources may vary, the stroke volume calculated from BEI measurements may have significant error.

17.6 Body Composition (Single-Frequency Measurement)

The percentage of body fat has been an important parameter in sports medicine, physical conditioning, weight loss programs, and predicting optimum body weight. To determine body composition, the body is configured as two parallel cylinders similar to the model described earlier. One cylinder is represented by fat and the other as fat-free body tissue. Since the resistivity of fat is much larger than muscle and other body fluids, the volume determined from the total body impedance measurement is assumed to represent the fat-free body volume. Studies have been conducted that calculate the fat-free body mass by determining the volume of the fat-free tissue cylinder using the impedance measured between the right hand and right foot, and using the subject's height as a measure of the cylinder's length with an empirical constant used to replace ρ [Lukaski, 1985]. Knowing the total weight and assuming body density factors, the percentage of body fat can be calculated as the difference between total weight and the weight of the fat-free tissue. Many studies have reported the correlation of the percentage of body fat calculated from BEI with other accepted standard techniques from approximately 0.88 to 0.98 [Schoeller and Kushner, 1989]. The physical model used for the equation development is a poor approximation to the actual body because it assumes a uniform cross-sectional area for the body between the hand the foot. Patterson [1989] pointed out the main problem with the technique: the measured impedance depends mostly on the characteristics of the arms and legs and not of the trunk. Therefore, determination of body fat with this method may often be inaccurate.

17.7 Impedance Spectroscopy

By measuring BEI over a range of frequencies (typically between 1 kHz to 1 MHz), the material properties of the tissues can be determined [Ackmann and Seitz, 1984]. Figure 17.6 shows the typical

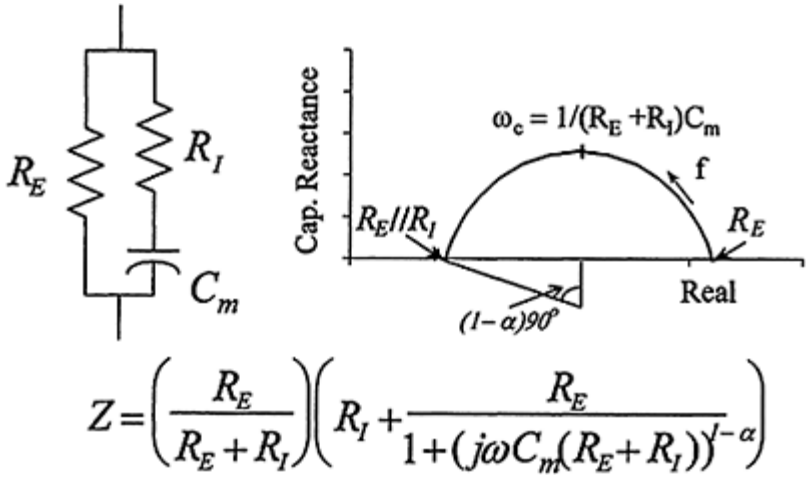


FIGURE 17.6 Typical impedance spectroscopy data, model, and the equation used to fit the data.

complex plane plot of the real and imaginary part of impedance and the model used to fit the data. R_E represents the extra-cellular space, R_I the intracellular space, and C_m the cell membrane. The parameter α is proportional to the angle of the suppression of the semicircle. It exists to account for the distribution of time constants in the tissue. At low frequencies, the current flows in the extracellular space and at high frequencies the current is capacitively passed across the cell membrane while the extra- and intracell spaces are in parallel.

Using these model parameters, studies have shown positive results in determining intra- and extracellular fluid volumes [Kanai et al., 1987], body fat [De Lorenzo et al., 1997], tissue ischemica [Cinca et al., 1997] and cancerous tissues [Jossient, 1998].

17.8 Summary

Electrical impedance instrumentation is not relatively costly, which has encouraged its possible application in many different areas. The impedance measurement is influenced by many different factors including geometry, tissue conductivity, and blood flow. Because of this complexity, it is difficult to reliably measure an isolated physiological parameter, which has been the principle factor limiting its use. The applications that are widely used in clinical medicine are apnea monitoring and the detection of venous thrombosis. The other applications described above will need more study before becoming a reliable and useful measurement.

Defining Terms

Apnea: A suspension of respiration.
Compliance: The volume change divided by the pressure change. The higher the compliance, the more easily the vessel will expand as pressure increases.
Plethysmography: The measurement of the volume change of an organ or body part.
Thrombosis: The formation of a thrombus.
Thrombus: A clot of blood formed within the heart or vessel.
Venogram: An X-ray image of the veins using an injected radiopaque contrast material.

References

Ackmann JJ, Seitz MA. 1984. Methods of complex impedance measurements in biologic tissue. *Crit Rev Biomed Eng* 11:281–311.

Anderson FA Jr. 1988. Impedance Plethysmography, in JG Webster (ed). *Encyclopedia of Medical Devices and Instrumentation*, Wiley, New York, pp 1632–1643.

Cinca J, Warran M, Carreno A, Tresanchez M. Armadans L, Gomez P, Soler-Soler J. 1997. Changes in myocardial electrical impedance induced by coronary artery occlusion in pigs with and without preconditioning. *Circulation* 96:3079–3086.

De Lorenzo A, Andreoi A, Matthie J, Withers P. 1997. Predicting body cell mass with bioimpedance by using theoretical methods: A technological review. *J Appl Physiol* 82(5):1542–1558.

Geddes LA, Baker LE. 1989. *Principles of Applied Biomedical Instrumentation—Third Edition*, pp 569–572, Wiley, New York.

Jossient J. 1998. The impedivity of freshly excised human breast tissue. *Physiol Meas* 19:61–75.

Kanai H, Haeno M, Sakamoto K. 1987. Electrical measurement of fluid distribution in human legs and arms. *Med Prog Technology* 12:159–170.

Lukaski HC, Johnson PE, Bolonchuk WW, Lykken GI. 1985. Assessment of fat-free mass using bioelectric impedance measurements of the human body. *Am J Clin Nutr* 41:810–817.

Mohapatra SN, Costeloe KL, Hill, DW. 1977. Blood resistivity and its implications for the calculations of cardiac output by the thoracic electrical impedance technique. *Intensive Care Medicine* 3:63.

Neuman MR. 1988. Neonatal monitoring, in JG Webster (ed). *Encyclopedia of Medical Devices and Instrumentation*, Wiley, New York, pp 2020–2034.

Nyboer J. 1970. *Electrical Impedance Plethysmography*, 2nd ed., Charles C Thomas, Springfield, IL.

Patterson RP. 1989. Fundamentals of Impedance Cardiography, *IEEE Engineering in Medicine and Biology Magazine* 8:35–38.

Patterson RP. 1989. Body Fluid Determinations Using Multiple Impedance Measurements, *IEEE Engineering in Medicine and Biology Magazine* 8:16–18.

Patterson R, Kubicek WG, Kinnen E, Witsoe D, Noren G. 1964. Development of an electrical impedance plethysmography system to monitor cardiac output. *Proc of the First Ann Rocky Mountain Bioengineering Symposium*, pp 56–71.

Patterson RP, Wang L, Raza SB. 1991. Impedance cardiography using band and regional electrodes in supine, sitting, and during exercise. *IEEE Trans on BME* 38:393–400.

Schoeller DA, Kushner RE 1989. Determination of Body Fluids by the Impedance Technique. *IEEE Engineering in Medicine and Biology Magazine* 8:19–21.

Wang L, Patterson RP. 1995. Multiple sources of the impedance cardiogram based on 3D finite difference human thorax models. *IEEE Trans Biomed Eng* 42:393–400.

Further Information

The book by Nyboer, *Electrical Impedance Plethysmography*, contains useful background information. *The Encyclopedia of Medical Devices and Instrumentation*, edited by J.G.Webster, and *Principles of Applied Biomedical Instrumentation*, by L.A.Geddes and L.E.Baker, give a more in-depth description of many applications and describe some usual measurements.

18
Respiration

Leslie A.Geddes
Purdue University

18.1 Lung Volumes

The amount of air flowing into and out of the lungs with each breath is called the tidal volume (TV). In a typical adult this amounts to about 500 mL during quiet breathing. The respiratory system is capable of moving much more air than the tidal volume. Starting at the *resting expiratory level* (REL in Fig. 18.1), it is possible to inhale a volume amounting to about seven times the tidal volume; this volume is called the *inspiratory capacity* (IC). A measure of the ability to inspire more than the tidal volume is the *inspiratory reserve volume* (IRV), which is also shown in Fig. 18.1. Starting from REL, it is possible to forcibly exhale a volume amounting to about twice the tidal volume; this volume is called the *expiratory reserve volume* (ERV). However, even with the most forcible expiration, it is not possible to exhale all the air from the lungs; a *residual volume* (RV) about equal to the expiratory reserve volume remains. The sum of the expiratory reserve volume and the residual volume is designated *the functional residual capacity* (FRC). The volume of air exhaled from a maximum inspiration to a maximum expiration is called the *vital capacity* (VC). The *total lung capacity* (TLC) is the total air within the lungs, i.e., that which can be moved in a vital-capacity maneuver plus the residual volume. All except the residual volume can be determined with a volume-measuring instrument such as a spirometer connected to the airway.

18.2 Pulmonary Function Tests

In addition to the static lung volumes just identified, there are several time-dependent volumes associated with the respiratory act. The *minute volume* (MV) is the volume of air per breath (tidal volume) multiplied by the respiratory rate (R), i.e., MV=(TV) R. It is obvious that the same minute volume can be produced by rapid shallow or slow deep breathing. However, the effectiveness is not the same, because not all the respiratory air participates in gas exchange, there being a dead space volume. Therefore the alveolar ventilation is the important quantity that is defined as the tidal volume (TV) minus the dead space (DS) multiplied by the respiratory rate R, i.e., alveolar ventilation=(TV–DS) R. In a normal adult subject, the dead space amounts to about 150 mL, or 2 mL/kg.

0-8493-1813-0/03/$0.00+$ 1.50

Dynamic Tests

Several timed respiratory volumes describe the ability of the respiratory system to move air. Among these are *forced vital capacity* (FVC), *forced expiratory volume* in *t* seconds (FEV_t), the *maximum ventilatory volume* (MVV), which was previously designated the *maximum breathing capacity* (MBC), and the *peak*

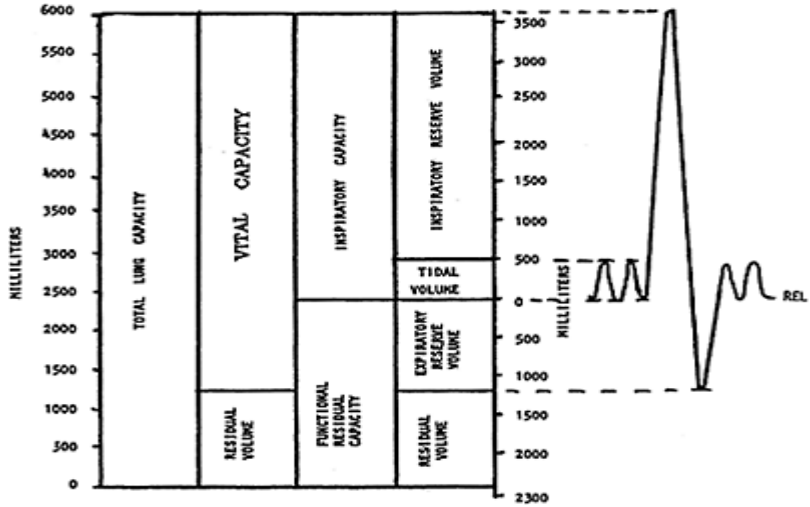


FIGURE 18.1 Lung volumes.

flow (PF). These quantities are measured with a spirometer without valves and CO_2 absorber or with a pneumotachograph coupled to an integrator.

Forced Vital Capacity

Forced vital capacity (FVC) is shown in Fig. 18.2 and is measured by taking the maximum inspiration and forcing all of the inspired air out as rapidly as possible. Table 18.1 presents normal values for males and females.

Forced Expiratory Volume

Forced expiratory volume in *t* seconds (FEV_t) is shown in Fig. 18.2, which identifies $FEV_{0.5}$ and $FEV_{1.0}$, and Table 18.1 presents normal values for $FEV_{1.0}$.

Maximum Voluntary Ventilation

Maximum voluntary ventilation (MVV) is the volume of air moved in 1 minute when breathing as deeply and rapidly as possible. The test is performed for 20 s and the volume scaled to a 1-min value; Table 18.1 presents normal values.

Peak Flow

Peak flow (PF) in L/min is the maximum flow velocity attainable during an FEV maneuver and represents the maximum slope of the expired volume-time curve (Fig. 18.2); typical normal values are shown in Table 18.1.

The Water-Sealed Spirometer

The water-sealed spirometer was the traditional device used to measure the volume of air moved in respiration. The Latin word *spirare* means to breathe. The most popular type of spirometer consists of a hollow cylinder closed at one end, inverted and suspended in an annular space filled with water to provide an air-tight seal. Figure 18.3 illustrates the method of suspending the counterbalanced cylinder (bell), which is free to move up and down to accommodate the volume of air under it. Movement of the bell, which is proportional to volume, is usually recorded by an inking pen applied to a graphic record

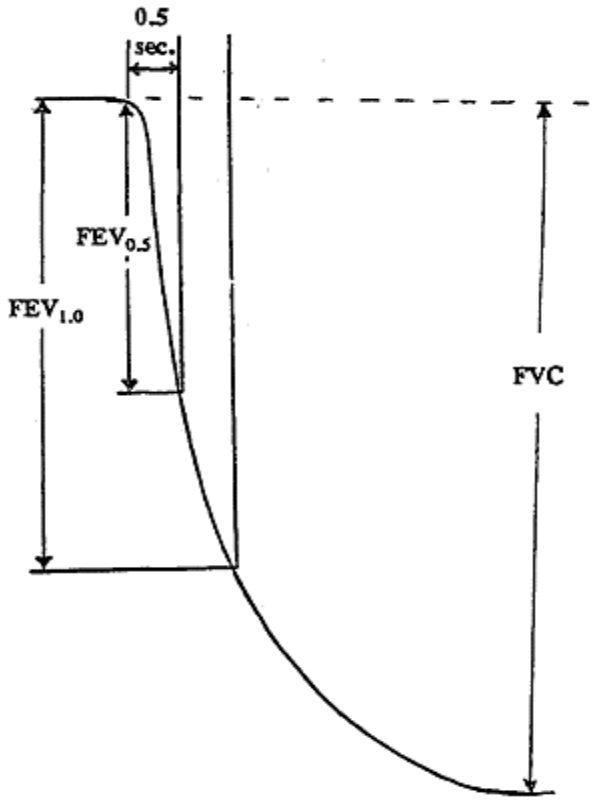


FIGURE 18.2 The measurement of timed forced expiratory volume (FEV_t) and forced vital capacity (FVC).

TABLE 18.1 Dynamic Volumes

Males	
	FVC (L)=0.133H—0.022A—3.60 (SEE—0.58)*
	FEV1 (L)=0.094H—0.028A—1.59 (SEE—0.52)*
	MVV (L/min)—3.39H—1.26A—21.4 (SEE—29)*
	PF (L/min)—(10.03–0.038A)H†
Females	
	FVC (L)=0.111H−0.015A−3.16 (SD=0.42)‡
	FEV1 (L)=0.068H—0.023A—0.92 (SD=0.37)‡
	MVV (L/min)—2.05H—0.57A—5.5 (SD=10.7)‡
	PF (L/min)=(7.44–0.0183A)H‡

H=height in inches, A=age in years, L=liters, L/min=liters per minute,
SEE=standard error of estimate, SD=standard deviation
*Kory, Callahan, Boren, Syner. 1961. *Am J Med* 30:243.
†Leiner, Abramowitz, Small, Stenby, Lewis. 1963. *Amer Rev Resp Dis* 88:644.
‡Lindall, Medina, Grismer. 1967. *Amer Rev Resp Dis* 95:1061.

that is caused to move with a constant speed. Below the cylinder, in the space that accommodates the volume of air, are inlet and outlet breathing tubes. At the end of one or both of these tubes is a check valve designed to maintain a unidirectional flow of air through the spirometer. Outside the spirometer the two breathing tubes are brought to a Y tube, which is connected to a mouthpiece. With a pinch clamp placed on the nose, inspiration diminishes the volume of air under the bell, which descends, causing the stylus to rise on the graphic record. Expiration produces the reverse effect. Thus, starting with the spirometer half-filled, quiet respiration causes the bell to rise and fall. By knowing the "bell factor," the volume of air moved per centimeter excursion of the bell, the volume change can be quantitated. Although a variety of flowmeters are now used to measure respiratory volumes, the spirometer with a CO_2 absorber is ideally suited to measure oxygen uptake.

Oxygen Uptake

A second and very important use for the water-filled spirometer is measurement of oxygen used per unit of time, designated the *O_2 uptake*. This measurement is accomplished by incorporating a soda-lime, carbon-dioxide absorber into the spirometer as shown in Fig. 18.4*a*. Soda-lime is a mixture of calcium hydroxide, sodium hydroxide, and silicates of sodium and calcium. The exhaled carbon dioxide combines with the soda-lime and becomes solid carbonates. A small amount of heat is liberated by this reaction.

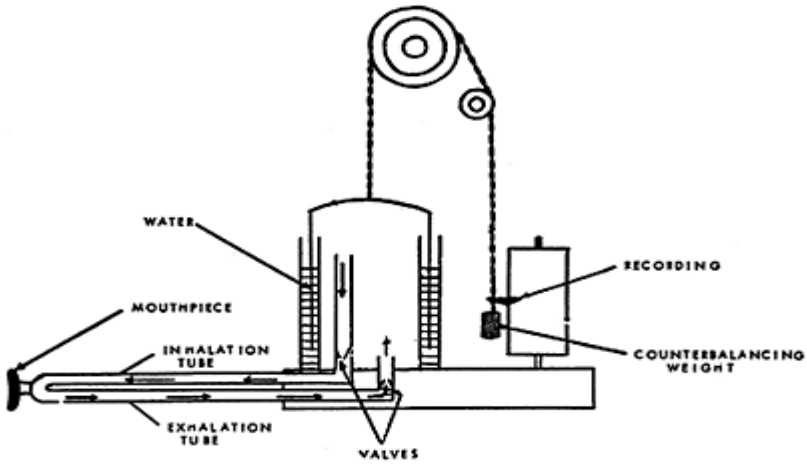


FIGURE 18.3 The simple spirometer.

Starting with a spirometer filled with oxygen and connected to a subject, respiration causes the bell to move up and down (indicating tidal volume) as shown in Fig. 18.5. With continued respiration the baseline of the recording rises, reflecting disappearance of oxygen from under the bell. By measuring the slope of the baseline on the spirogram, the volume of oxygen consumed per minute can be determined. Figure 18.5 presents a typical example along with calculation.

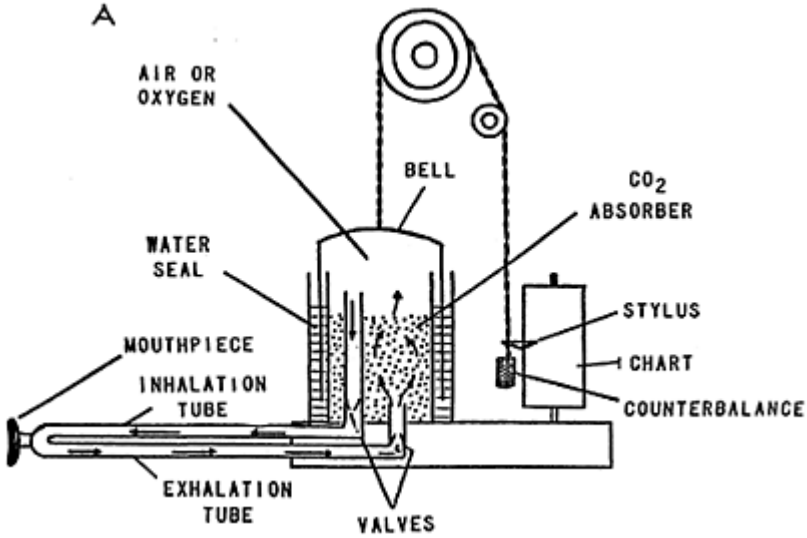
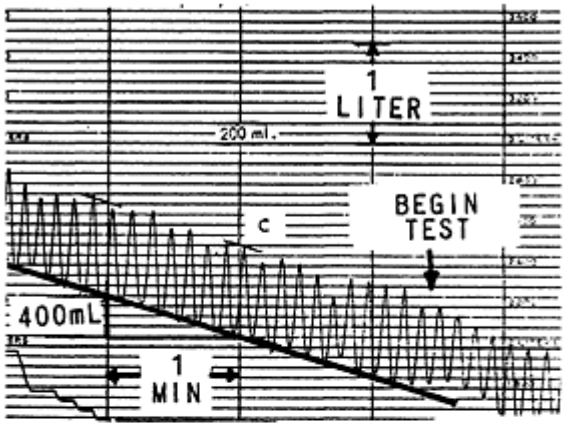


FIGURE 18.4 The spirometer with CO_2 absorber (Fig. 18.5) and a record of oxygen uptake.



$$V_{BTPS} = V_{MEAS} \times F$$

$$F = \frac{273+37}{273+T} \times \frac{P_B - P_{H_2O}}{P_B - 47}$$

V_{MEAS} = 400mL @ 26°C = T
(P_B=750 mmHg)

$$F = \frac{273+37}{273+26} \times \frac{750-25.2}{750-47}$$

$$= 1.069$$

$$V_{BTPS} = 400 \times 1.069$$

$$= 427.6 \text{ mL}$$

FIGURE 18.5 Oxygen consumption.

The Dry Spirometer

The water-sealed spirometer was the most popular device for measuring the volumes of respiratory gases; however, it is not without its inconveniences. The presence of water causes corrosion of the metal parts. Maintenance is required to keep the device in good working order over prolonged periods. To eliminate these problems, manufacturers have developed dry spirometers. The most common type employs a collapsible rubber or plastic bellows, the expansion of which is recorded during breathing. The earlier rubber models had a decidedly undesirable characteristic that caused their abandonment. When the bellows was in its midposition, the resistance to breathing was a minimum; when fully collapsed, it imposed a slight negative resistance; and when fully extended it imposed a slight positive resistance to breathing. Newer units with compliant plastic bellows minimize this defect.

The Pneumotachograph

The pneumotachograph is a device that is placed directly in the airway to measure the velocity of air flow. The volume per breath is therefore the integral of the velocity-time record during inspiration or expiration. Planimetric integration of the record, or electronic integration of the velocity-time signal, yields the tidal volume. Although tidal volume is perhaps more easily recorded with the spirometer, the dynamics of respiration are better displayed by the pneumotachograph, which offers less resistance to the air stream and exhibits a much shorter response time—so short in most instruments that cardiac impulses are often clearly identifiable in the velocity-time record.

If a specially designed resistor is placed in a tube in which the respiratory gases flow, a pressure drop will appear across it. Below the point of turbulent flow, the pressure drop is linearly related to air-flow velocity. The resistance may consist of a wire screen or a series of capillary tubes; Fig. 18.6 illustrates both

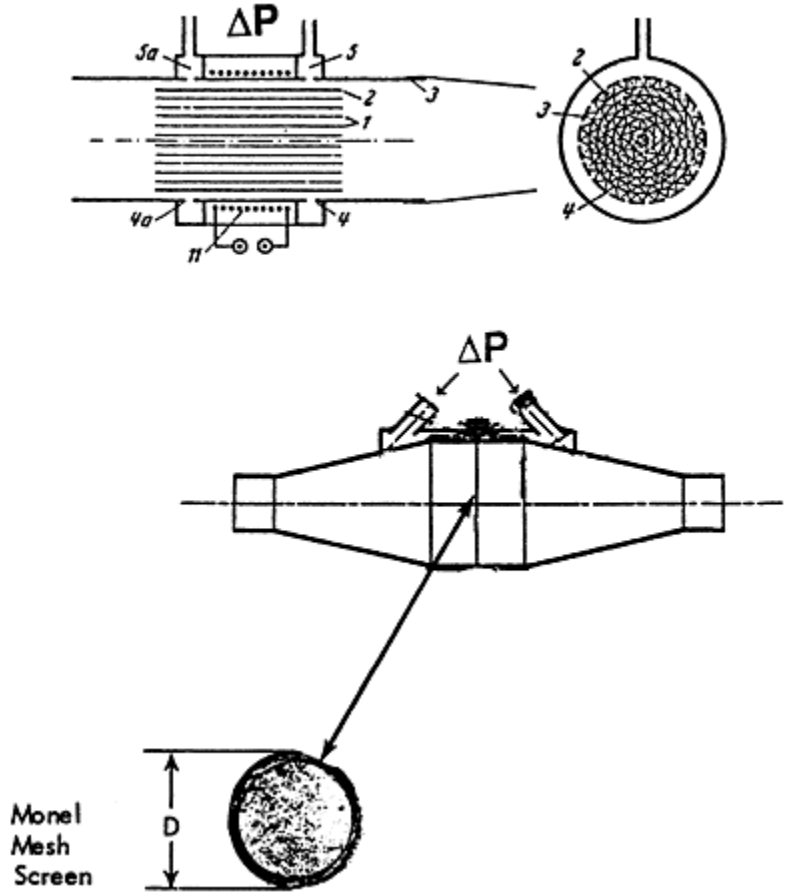


FIGURE 18.6 Pneumotachographs.

types. Detection and recording of this pressure differential constitutes a pneumotachogram; Fig. 18.7 presents a typical air-velocity record, along with the spirogram, which is the integral of the flow signal. The small-amplitude artifacts in the pneumotachogram are cardiac impulses.

For human application, linear flow rates up to 200 L/min should be recordable with fidelity. The resistance to breathing depends upon the flow rate, and it is difficult to establish an upper limit of tolerable resistance. Silverman and Whittenberger [1950] stated that a resistance of 6 mm H_2O is perceptible to human subjects. Many of the high-fidelity pneumotachographs offer 5 to 10 mm H_2O resistance at 100 and 200 L/min. It would appear that such resistances are acceptable in practice.

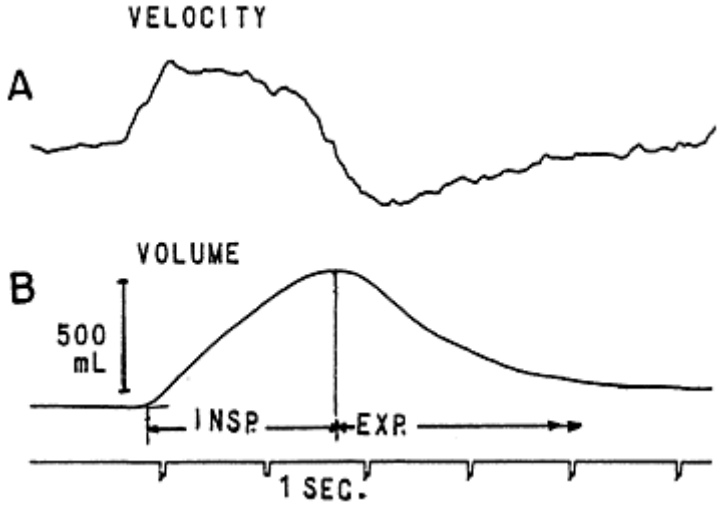


FIGURE 18.7 Velocity (A) and volume changes (B) during normal, quiet breathing; B is the integral of A.

Response times of 15 to 40 ms seem to be currently in use. Fry and coworkers [1957] analyzed the dynamic characteristics of three types of commercially available, differential-pressure pneumotachographs that employed concentric cylinders, screen mesh, and parallel plates for the air resistors. Using a high-quality, differential-pressure transducer with each, they measured total flow resistance ranging from 5 to 15 cm H_2O. Frequency response curves taken on one model showed fairly uniform response to 40 Hz; the second model showed a slight increase in response at 50 Hz, and the third exhibited a slight drop in response at this frequency.

The Nitrogen-Washout Method for Measuring FRC

The *functional residual capacity* (FRC) and the *residual volume* (RV) are the only lung compartments that cannot be measured with a volume-measuring device. Measuring these requires use of the nitrogen analyzer and application of the dilution method.

Because nitrogen does not participate in respiration, it can be called a *diluent*. Inspired and expired air contain about 80% nitrogen. Between breaths, the FRC of the lungs contains the same concentration of nitrogen as in environmental air, i.e., 80%. By causing a subject to inspire from a spirometer filled with 100% oxygen and to exhale into a second collecting spirometer, all the nitrogen in the FRC is replaced by oxygen, i.e., the nitrogen is "washed out" into the second spirometer. Measurement of the concentration of nitrogen in the collecting spirometer, along with a knowledge of its volume, permits calculation of the amount of nitrogen originally in the functional residual capacity and hence allows calculation of the FRC, as now will be shown.

Figure 18.8 illustrates the arrangement of equipment for the nitrogen-washout test. Note that two check values, (I, EX) are on both sides of the subject's breathing tube, and the nitrogen meter is connected to the mouthpiece. Valve V is used to switched the subject from breathing environmental air to the measuring system. The left-hand spirometer contains 100% oxygen, which is inhaled by the subject via valve I. Of course, a nose clip must be applied so that all the respired gases flow through the breathing tube connected to the mouthpiece. It is in this tube that the sampling inlet for the nitrogen analyzer is located. Starting at the resting expiratory level, inhalation of pure oxygen causes the nitrogen analyzer to indicate zero. Expiration closes valve I and opens valve EX. The first expired breath contains nitrogen derived from the FRC (diluted by the oxygen that was inspired); the nitrogen analyzer indicates this percentage. The exhaled gases are collected in the right-hand spirometer. The collecting spirometer and all the interconnecting tubing was first flushed with oxygen to eliminate all nitrogen. This simple procedure eliminates the need for applying corrections and facilitates calculation of the FRC. With

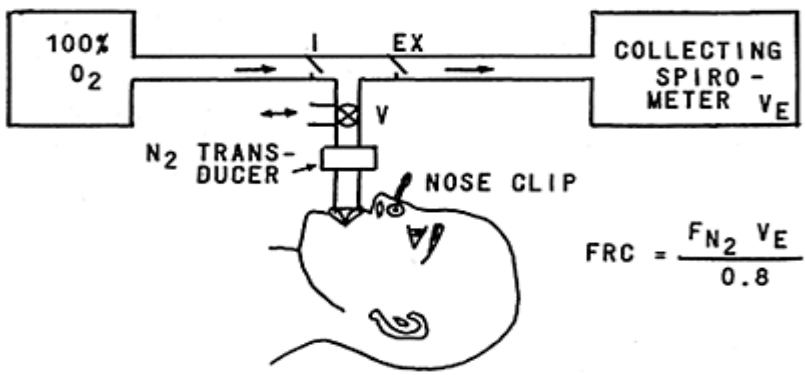


FIGURE 18.8 Arrangement of equipment for the nitrogen-washout technique. Valve V allows the subject

to breathe room air until the test is started. The test is started by operating valve V at the end of a normal breath, i.e. the subject starts breathing 100% O_2 through the inspiratory valve (I) and exhales the N_2 and O_2 mixture into a collecting spirometer via the expiratory valve EX.

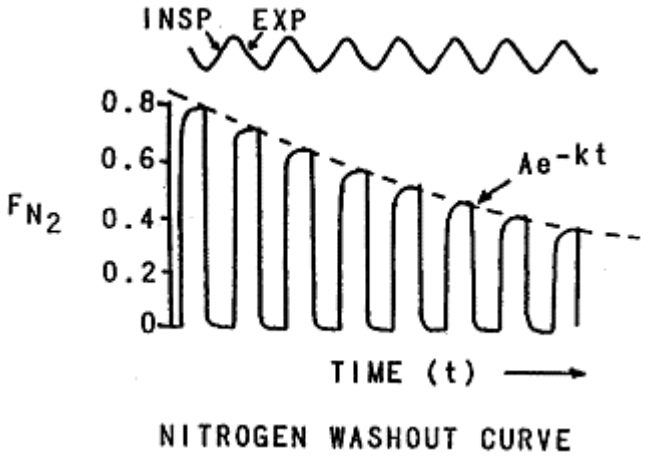


FIGURE 18.9 The nitrogen-washout out curve.

continued breathing, the nitrogen analyzer indicates less and less nitrogen because it is being washed out of the FRC and is replaced by oxygen. Figure 18.9 presents a typical record of the diminishing concentration of expired nitrogen throughout the test. In most laboratories, the test is continued until the concentration of nitrogen falls to about 1%. The nitrogen analyzer output permits identification of this concentration. In normal subjects, virtually all the nitrogen can be washed out of the FRC in about 5 minutes.

If the peaks on the nitrogen-washout record are joined, a smooth exponential decay curve is obtained in normal subjects. A semilog of N_2 versus time provides a straight line. In subjects with trapped air, or poorly ventilated alveoli, the nitrogen-washout curve consists of several exponentials as the multiple poorly ventilated regions give up their nitrogen. In such subjects, the time taken to wash out all the nitrogen usually exceeds 10 min. Thus, the nitrogen concentration-time curve provides useful diagnostic information on ventilation of the alveoli.

If it is assumed that all the collected (washed-out) nitrogen was uniformly distributed within the lungs, it is easy to calculate the FRC. If the environmental air contains 80% nitrogen, then the volume of nitrogen in the functional residual capacity is 0.8 (FRC). Because the volume of expired gas in the collecting spirometer is known, it is merely

necessary to determine the concentration of nitrogen in this volume. To do so requires admitting some of this gas to the inlet valve of the nitrogen analyzer. Note that this concentration of nitrogen *(F_{N2})* exists in a volume that includes the volume of air expired *(V_E)* plus the original volume of oxygen in the collecting spirometer (V_o) at the start of the test and the volume of the tubing *(V_t)* leading from the expiratory collecting valve. It is therefore advisable to start with an empty collecting spirometer (V_o=0). Usually the tubing volume *(V_t)* is negligible with respect to the volume of expired gas collected in a typical washout test. In this situation the volume of nitrogen collected is $V_E F_{N2}$, where F_{N2} is the fraction of nitrogen within the collected gas. Thus, 0.80 (FRC)=F_{N2} *(V_E)*. Therefore

$$\mathrm{FRC} = \frac{F_{N2}\, V_E}{0.80}.$$

It is important to note that the value for FRC so obtained is at ambient temperature and pressure and is saturated with water vapor (ATPS). In respiratory studies, this value is converted to body temperature and saturated with water vapor (BTPS).

In the example shown in Fig. 18.9, the washout to 1% took about 44 breaths. With a breathing rate of 12/min, the washout time was 220 s. The volume collected *(V_E)* was 22 L and the concentration of nitrogen in this volume was 0.085 *(F_{N2});* therefore

$$\mathrm{FRC} = \frac{0.085 \times 22000}{0.80} = 2337\,\mathrm{mL}.$$

18.3 Physiologic Dead Space

The volume of ventilated lung that does not participate in gas exchange is the physiologic dead space *(V_d)*. It is obvious that the physiologic dead space includes anatomic dead space, as well as the volume of any alveoli that are not perfused. In the lung, there are theoretically four types of alveoli, as shown in Fig. 18.10. The normal alveolus (A) is both ventilated and perfused with blood. There are alveoli that are ventilated but not perfused (B); such alveoli contribute significantly to the physiologic dead space. There are alveoli that are not ventilated but perfused (C); such alveoli do not provide the exchange of respiratory gases. Finally, there are alveoli that are both poorly ventilated and poorly perfused (D); such alveoli contain high CO_2 and N_2 and low O_2. These alveoli are the last to expel their CO_2 and N_2 in washout tests.

Measurement of physiologic dead space is based on the assumption that there is almost complete equilibrium between alveolar pCO_2 and pulmonary capillary blood. Therefore, the arterial pCO_2 represents mean alveolar pCO_2 over many breaths when an arterial blood sample is drawn for analysis of pCO_2. The Bohr equation for physiologic dead space is

$$V_d = \left[\frac{\mathrm{PaCO_2} - \mathrm{pECO_2}}{\mathrm{PaCO_2}}\right] V_E.$$

In this expression, $paCO_2$ is the partial pressure in the arterial blood sample, which is withdrawn slowly during the test; $pECO_2$ is the partial pressure of CO_2 in the volume of expired air; V_E is the volume of expired air per breath (tidal volume).

In a typical test, the subject would breathe in room air and exhale into a collapsed (Douglas) bag. The test is continued for 3 min or more, and the number of breaths is counted in that period. An arterial blood sample is withdrawn during the collection period. The pCO_2 in the expired gas is measured, and then the volume of expired gas is measured by causing it to flow into a spirometer or flowmeter by collapsing the collecting bag.

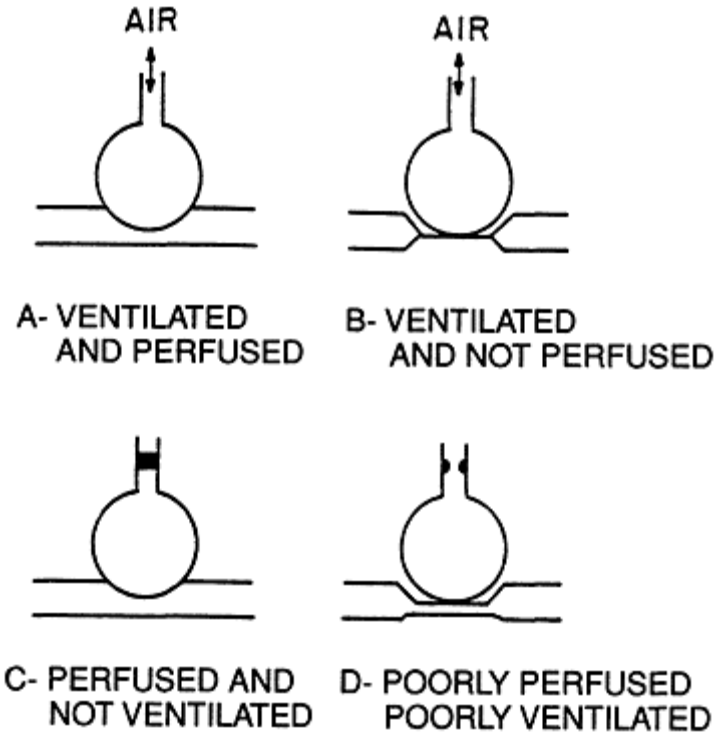


FIGURE 18.10 The four types of alveoli.

In a typical 3-min test, the collected volume is 33 L, and the pCO_2 in the expired gas is 14.5 mmHg. During the test, the pCO_2 in the arterial blood sample was 40 mmHg. The number of breaths was 60; therefore, the average tidal volume was 33000/60=550 mL. The physiologic dead space (V_d) is:

$$V_d = \left[\frac{40-14.5}{40}\right] 550 = 350\,\text{mL}.$$

It is obvious that an elevated physiological dead space indicates lung tissue that is not perfused with blood.

References

Fry DI, Hyatt RE, McCall CB. 1957. Evaluation of three types or respiratory flowmeters. *Appl Physiol* 10:210.

Silverman L, Whittenberger J. 1950. Clinical pneumatachograph. *Methods Med Res* 2:104.

19
Clinical Laboratory: Separation and Spectral Methods

Richard L.Roa
Baylor University Medical Center

The purpose of the clinical laboratory is to analyze body fluids and tissues for specific substances of interest and to report the results in a form that is of value to clinicians in the diagnosis and treatment of disease. A large range of tests has been developed to achieve this purpose. Four terms commonly used to describe tests are *accuracy*, *precision*, *sensitivity*, and *specificity*. An accurate test, on average, yields true values. Precision is the ability of a test to produce identical results upon repeated trials. Sensitivity is a measure of how small an amount of substance can be measured. Specificity is the degree to which a test measures the substance of interest without being affected by other substances that may be present in greater amounts.

The first step in many laboratory tests is to separate the material of interest from other substances. This may be accomplished through extraction, filtration, and centrifugation. Another step is derivatization, in which the substance of interest is chemically altered through addition of reagents to change it into a substance that is easily measured. For example, one method for measuring glucose is to add o-toluidine which, under proper conditions, forms a green-colored solution with an absorption maximum at 630 nm. Separation and derivatization both improve the specificity required of good tests.

19.1 Separation Methods

Centrifuges are used to separate materials on the basis of their relative densities. The most common use in the laboratory is the separation of cells and platelets from the liquid part of the blood. This requires a relative centrifugal force (RCF) of roughly 1000 *g* (1000 times the force of gravity) for a period of 10 minutes. Relative centrifugal force is a function of the speed of rotation and the distance of the sample from the center of rotation as stated in Eq. (19.1):

$$\mathrm{RCF} = \left(1.12 \times 10^{-5}\right) r \left(\mathrm{rpm}\right)^2, \tag{19.1}$$

where RCF=relative centrifugal force in *g*, and *r*=radius in cm.

Some mixtures require higher g-loads in order to achieve separation in a reasonable period of time. Special rotors contain the sample tubes inside a smooth container, which

0-8493-1813-0/03/$0.00+$1.50

minimizes air resistance to allow faster rotational speeds. Refrigerated units maintain the samples at a cool temperature throughout long high-speed runs that could lead to sample heating due to air friction on the rotor. Ultracentrifuges operate at speeds on the order of 100,000 rpm and provide relative centrifugal forces of up to 600,000 *g*. These usually require vacuum pumps to remove the air that would otherwise retard the rotation and heat the rotor.

19.2 Chromatographic Separations

Chromatographic separations depend upon the different rates at which various substances moving in a stream (mobile phase) are retarded by a stationary material (stationary phase) as they pass over it. The mobile phase can be a volatilized sample transported by an inert carrier gas such as helium or a liquid transported by an organic solvent such as acetone. Stationary phases are quite diverse depending upon the separation being made, but most are contained within a long, thin tube (column). Liquid stationary phases may be used by coating them onto inert packing materials. When a sample is introduced into a Chromatographic column, it is carried through it by the mobile phase. As it passes through the column, the substances that have greater affinity for the stationary phase fall behind those with less affinity. The separated substances may be detected as individual peaks by a suitable detector placed at the end of the Chromatographic column.

19.3 Gas Chromatography

The most common instrumental Chromatographic method used in the clinical laboratory is the gas-liquid chromatograph. In this system the mobile phase is a gas, and the stationary phase is a liquid coated onto either an inert support material, in the case of a packed column, or the inner walls of a very thin tube, in the case of a capillary column. Capillary columns have the greatest resolving power but cannot handle large sample quantities. The sample is injected into a small heated chamber at the beginning of the column, where it is volatilized if it is not already a gaseous sample. The sample is then carried through the column by an inert carrier gas, typically helium or nitrogen. The column is completely housed within an oven. Many gas chromatographs allow for the oven temperature to be programmed to slowly increase for a set time after the sample injection is made. This produces peaks that are spread more uniformly over time.

Four detection methods commonly used with gas chromatography are thermal conductivity, flame ionization, nitrogen/phosphorous, and mass spectrometry. The thermal conductivity detector takes advantage of variations in thermal conductivity between the carrier gas and the gas being measured. A heated filament immersed in the gas leaving the Chromatographic column is part of a Wheatstone bridge circuit. Small variations in the conductivity of the gas cause changes in the resistance of the filament, which are recorded. The flame ionization detector measures the current between two plates with a voltage applied between them. When an organic material appears in the flame, ions that contribute to the current are formed. The NP detector, or

nitrogen/phosphorous detector, is a modified flame ionization detector (see Fig. 19.1) that is particularly sensitive to nitrogen- and phosphorous-containing compounds.

Mass spectrometry (MS) provides excellent sensitivity and selectivity. The concept behind these devices is that the volatilized sample molecules are broken into ionized fragments that are then passed through a mass analyzer that separates the fragments according to their mass/charge *(m/z)* ratios. A mass spectrum, which is a plot of the relative abundance of the various fragments versus m/z, is produced. The mass spectrum is characteristic of the molecule sampled. The mass analyzer most commonly used with gas

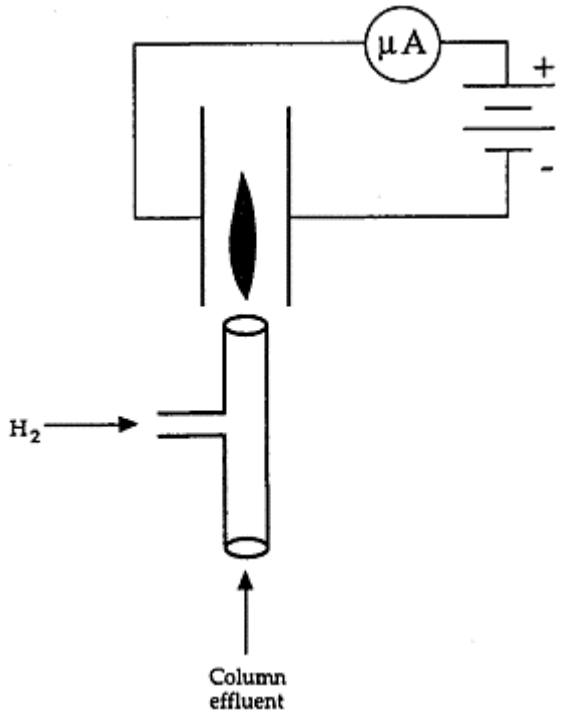


FIGURE 19.1 Flame ionization detector. Organic compounds in the column effluent are ionized in the flame, producing a current proportional to the amount of the compound present.

chromatographs is the quadrupole detector, which consists of four rods that have dc and RF voltages applied to them. The *m/z* spectrum can be scanned by appropriate changes in the applied voltages. The detector operates in a manner similar to that of a photomultiplier tube except that the collision of the charged particles with the cathode

begins the electron cascade, resulting in a measurable electric pulse for each charged particle captured. The MS must operate in a high vacuum, which requires good pumps and a porous barrier between the GC and MS that limits the amount of carrier gas entering the MS.

19.4 High-Performance Liquid Chromatography

In liquid chromatography, the mobile phase is liquid. High-performance liquid chromatography (HPLC) refers to systems that obtain excellent resolution in a reasonable time by forcing the mobile phase at high pressure through a long, thin column. The most common pumps used are pistons driven by asymmetrical cams. By using two such pumps in parallel and operating out of phase, pressure fluctuations can be minimized. Typical pressures are 350 to 1500 psi, though the pressure may be as high as 10,000 psi. Flow rates are in the 1 to 10 mL/min range.

A common method for placing a sample onto the column is with a loop injector, consisting of a loop of tubing that is filled with the sample. By a rotation of the loop, it is brought in series with the column, and the sample is carried onto the column. A UV/visible spectrophotometer is often used as a detector for this method. A mercury arc lamp with the 254-nm emission isolated is useful for detection of aromatic compounds, while diode array detectors allow a complete spectrum from 190 nm to 600 nm in 10 msec. This provides for detection and identification of compounds as they come off the column. Fluorescent, electrochemical, and mass analyzer detectors are also used.

19.5 Basis for Spectral Methods

Spectral methods rely on the absorption or emission of electromagnetic radiation by the sample of interest. Electromagnetic radiation is often described in terms of frequency or wavelength. Wavelengths are those obtained in a vacuum and may be calculated with the formula,

$$\lambda = c/\upsilon, \tag{19.2}$$

where λ,=wavelength in meters

c=speed of light in vacuum (3×10^8 m/s)

v=frequency in Hz.

The frequency range of interest for most clinical laboratory work consists of the visible (390–780 nm) and the ultraviolet or UV (180–390 nm) ranges. Many substances absorb different wavelengths preferentially. When this occurs in the visible region, they are colored. In general, the color of a substance is the complement of the color it absorbs, e.g., absorption in the blue produces a yellow color. For a given wavelength or bandwidth, transmittance is defined as

$$T = \frac{I_t}{I_i}, \tag{19.3}$$

where T=transmittance ratio (often expressed as %)
I_i=incident light intensity
I_t=transmitted light intensity.
Absorbance is defined as

$$A = \log_{10} 1/T . \tag{19.4}$$

Under suitable conditions, the absorbance of a solution with an absorbing compound dissolved in it is proportional to the concentration of that compound as well as the path length of light through it. This relationship is expressed by Beer's law:

$$A = abc , \tag{19.5}$$

where A=absorbance
a=a constant
b=path length
c=concentration.

A number of situations may cause deviations from Beer's law, such as high concentration or mixtures of compounds that absorb at the wavelength of interest. From an instrumental standpoint, the primary causes are stray light and excessive spectral bandwidth. Stray light refers to any light reaching the detector other than light from the desired pass-band that has passed through sample. Sources of stray light may include room light leaking into the detection chamber, scatter from the cuvette, and undesired fluorescence.

A typical spectrophotometer consists of a light source, some form of wavelength selection, and a detector for measuring the light transmitted through the samples. There is no single light source that covers the entire visible and UV spectrum. The source most commonly used for the visible part of the spectrum is the tungsten-halogen lamp, which provides continuous radiation over the range of 360 to 950 nm. The deuterium lamp has become the standard for much UV work. It covers the range from 220 to 360 nm. Instruments that cover the entire UV/visible range use both lamps with a means for switching from one lamp to the other at a wavelength of approximately 360 nm (Fig. 19.2).

Wavelength selection is accomplished with filters, prisms, and diffraction gratings. Specially designed interference filters can provide bandwidths as small as 5 nm. These are useful for instruments that do not need to scan a range of wavelengths. Prisms produce a nonlinear dispersion of wavelengths with the longer wavelengths closer together than the shorter ones. Since the light must pass through the prism material, they must be made of quartz for UV work. Diffraction gratings are surfaces with 1000 to 3000

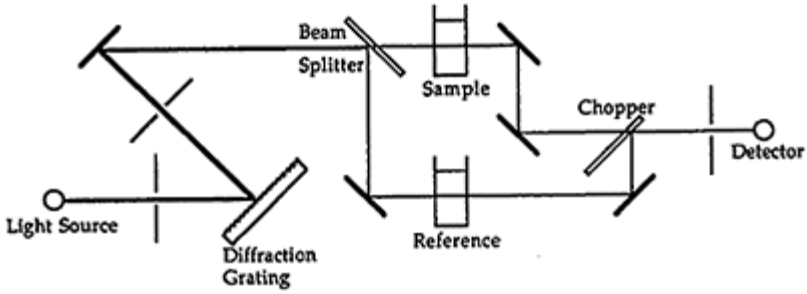


FIGURE 19.2 Dual-beam spectrophotometer. The diffraction grating is rotated to select the desired wavelength. The beam splitter consists of a half-silvered mirror that passes half the light while reflecting the other half. A rotating mirror with cut-out sections (chopper) alternately directs one beam and then the other to the detector.

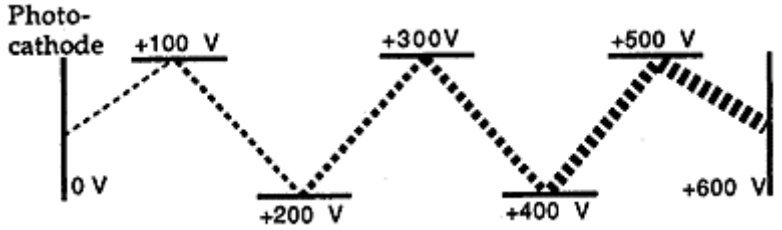


FIGURE 19.3 Photomultiplier tube. Incident photons cause the photocathode to emit electrons that collide with the first dynode, which emits additional electrons. Multiple dynodes provide sufficient gain to produce an easily measurable electric pulse from a single photon.

grooves/mm cut into them. They may be transmissive or reflective; the reflective ones are more popular since there is no attenuation of light by the material. They produce a linear dispersion. By proper selection of slit widths, pass bands of 0.1 nm are commonly achieved.

The most common detector is the photomultiplier tube, which consists of a photosensitive cathode that emits electrons in proportion to the intensity of light striking it (Fig. 19.3). A series of 10–15 dynodes, each at 50 to 100 volts greater potential than the preceding one, produce an electron amplification of 4 to 6 per stage. Overall gains are typically a million or more. Photomultiplier tubes respond quickly and cover the entire spectral range. They require a high voltage supply and can be damaged if exposed to room light while the high voltage is applied.

19.6 Fluorometry

Certain molecules absorb a photon's energy and then emit a photon with less energy (longer wavelength). When the reemission occurs in less than 10^{-8} s, the process is known as *fluorescence*. This physical process provides the means for assays which are 10 to 100 times as sensitive as those based on absorption measurements. This increase in sensitivity is largely because the light measured is all from the sample of interest. A dim light is easily measured against a black background, while it may be lost if added to an already bright background.

Fluorometers and spectrofluorometers are very similar to photometers and spectrophotometers but with two major differences. Fluorometers and spectrofluorometers use two monochrometers, one for excitation light and one for emitted light. By proper selection of the bandpass regions, all the light used to excite the sample can be blocked from the detector, assuring that the detector sees only fluorescence. The other difference is that the detector is aligned off-axis, commonly at 90°, from the excitation source. At this angle, scatter is minimal, which helps ensure a dark background for the measured fluorescence. Some spectrofluorometers use polarization filters both on the input and output light beams, which allows for fluorescence polarization studies (Fig. 19.4). An intense light source in the visible-to-UV range is

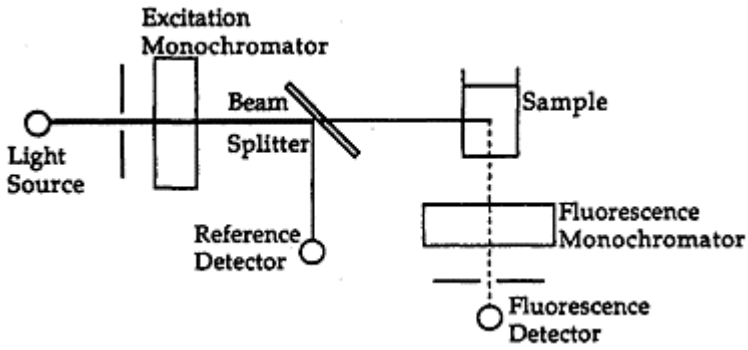


FIGURE 19.4 Spectrofluorometer. Fluorescence methods can be extremely sensitive to the low background interference. Since the

detector is off-axis from the incident light and a second monochromator blocks light of wavelengths illuminating the sample, virtually no signal reaches the detector other than the desired fluorescence.

desirable. A common source is the xenon or mercury arc lamps, which provide a continuum of radiation over this range.

19.7 Flame Photometry

Flame photometry is used to measure sodium, potassium, and lithium in body fluids. When these elements are heated in a flame they emit characteristic wavelengths of light. The major emission lines are 589 nm (yellow) for sodium, 767 nm (violet) for potassium, and 671 nm (red) for lithium. An atomizer introduces a fine mist of the sample into a flame. For routine laboratory use, a propane and compressed air flame is adequate. High-quality interference filters with narrow passbands are often used to isolate the major emission lines. The narrow bandpass is necessary to maximize the signal-to-noise ratio. Since it is impossible to maintain stable aspiration, atomization, and flame characteristics, it is necessary to use an internal standard of known concentration while making measurements of unknowns. In this way the ratio of the unknown sample's emission to the internal standard's emission remains stable even as the total signal fluctuates. An internal standard is usually an element that is found in very low concentration in the sample fluid. By adding a high concentration of this element to the sample, its concentration can be known to a high degree of accuracy. Lithium, potassium, and cesium all may be used as internal standards depending upon the particular assay being conducted.

19.8 Atomic Absorption Spectroscopy

Atomic absorption spectroscopy is based on the fact that just as metal elements have unique emission lines, they have identical absorption lines when in a gaseous or dissociated state. The atomic absorption spectrometer takes advantage of these physical characteristics in a clever manner, producing an instrument with approximately 100 times the sensitivity of a flame photometer for similar elements. The sample is aspirated into a flame, where the majority of the atoms of the element being measured remain in the ground state, where they are capable of absorbing light at their characteristic wavelengths. An intense source of exactly these wavelengths is produced by a hollow cathode lamp. These lamps are constructed so that the cathode is made from the element to be measured, and the lamps are filled with a low pressure of argon or neon gas. When a current is passed through the lamp, metal atoms are sputtered off the cathode and

collide with the argon or neon in the tube, producing emission of the characteristic wavelengths. A monochromator and photodetector complete the system.

Light reaching the detector is a combination of that which is emitted by the sample (undesirable) and light from the hollow cathode lamp that was not absorbed by the sample in the flame (desirable). By pulsing the light from the lamp either by directly pulsing the lamp or with a chopper, and using a detector that is sensitive to ac signals and insensitive to dc signals, the undesirable emission signal is eliminated. Each element to be measured requires a lamp with that element present in the cathode. Multielement

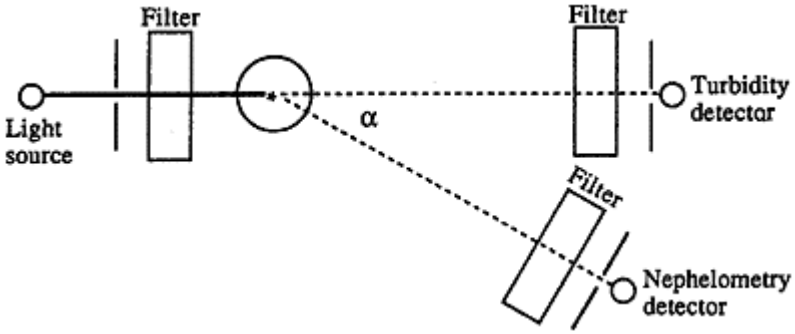


FIGURE 19.5 Nephelometer. Light scattered by large molecules is measured at an angle α away from the axis of incident light. The filters select the wavelength range desired and block undesired fluorescence. When α=0, the technique is known as turbidimetry.

lamps have been developed to minimize the number of lamps required. Atomic absorption spectrophotometers may be either single beam or double beam; the double-beam instruments have greater stability.

There are various nameless methods for atomic absorption spectroscopy in which the burner is replaced with a method for vaporizing the element of interest without a flame. The graphite furnace, which heats the sample to 2700°, consists of a hollow graphite tube which is heated by passing a large current through it. The sample is placed within the tube, and the light beam is passed through it while the sample is heated.

19.9 Turbidimetry and Nephelometry

Light scattering by particles in solution is directly proportional to both concentration and molecular weight of the particles. For small molecules the scattering is insignificant, but for proteins, immunoglobulins, immune complexes, and other large particles, light

scattering can be an effective method for the detection and measurement of particle concentration. For a given wavelength λ of light and particle size *d*, scattering is described as Raleigh ($d<\lambda/10$), Raleigh-Debye ($d\approx\lambda$), or Mie ($d>10\lambda$). For particles that are small compared to the wavelength, the scattering is equal in all directions. However, as the particle size becomes larger than the wavelength of light, it becomes preferentially scattered in the forward direction. Light-scattering techniques are widely used to detect the formation of antigen-antibody complexes in immunoassays.

When light scattering is measured by the attenuation of a beam of light through a solution, it is called *turbidimetry*. This is essentially the same as absorption measurements with a photometer except that a large passband is acceptable. When maximum sensitivity is required a different method is used—direct measurement of the scattered light with a detector placed at an angle to the central beam. This method is called *nephelometry*. A typical nephelometer will have a light source, filter, sample cuvette, and detector set at an angle to the incident beam (Fig. 19.5).

Defining Terms

Accuracy: The degree to which the average value of repeated measurements approximate the true value being measured.
Fluorescence: Emission of light by an atom or molecule following absorption of a photon by greater energy. Emission normally occurs within 10^{-8} of absorption.
Nephelometry: Measurement of the amount of light scattered by particles suspended in a fluid.
Precision: A measure of test reproducibility.
Sensitivity: A measure of how small an amount or concentration of an analyte can be detected.
Specificity: A measure of how well a test detects the intended analyte without being "fooled" by other substances in the sample.
Turbidimetry: Measurement of the attenuation of a light beam due to light lost to scattering by particles suspended in a fluid.

References

Burtis CA, Ashwood ER (eds). 1994. *Tietz Textbook of Clinical Chemistry,* 2d ed, Philadelphia, Saunders.
Hicks MR, Haven MC, Schenken JR, et al. (eds). 1987. *Laboratory Instrumentation,* 3d ed, Philadelphia: Lippincott.
Kaplan LA, Pesce AJ (eds). 1989. *Clinical Chemistry: Theory, Analysis, and Correlation,* 2d ed, St. Louis, Mosby.
Tietz NW (ed). 1987. *Fundamentals of Clinical Chemistry,* 3d ed, Philadelphia, Saunders.
Ward JM, Lehmann CA, Leiken AM. 1994. *Clinical Laboratory Instrumentation and Automation: Principles, Applications, and Selection,* Philadelphia, Saunders.

20
Clinical Laboratory: Nonspectral Methods and Automation

Richard L.Roa
Baylor University Medical Center

20.1 Particle Counting and Identification

The Coulter principle was the first major advance in automating blood cell counts. The cells to be counted are drawn through a small aperture between two fluid compartments, and the electric impedance between the two compartments is monitored (see Fig. 20.1). As cells pass through the aperture, the impedance increases in proportion to the volume of the cell, allowing large numbers of cells to be counted and sized rapidly. Red cells are counted by pulling diluted blood through the aperture. Since red cells greatly outnumber white cells, the contribution of white cells to the red cell count is usually neglected. White cells are counted by first destroying the red cells and using a more concentrated sample.

Modern cell counters using the Coulter principle often use *hydrodynamic focusing* to improve the performance of the instrument. A sheath fluid is introduced that flows along the outside of a channel with the sample stream inside it. By maintaining laminar flow conditions and narrowing the channel, the sample stream is focused into a very thin column with the cells in single file. This eliminates problems with cells flowing along the side of the aperture or sticking to it and minimizes problems with having more than one cell in the aperture at a time.

Flow cytometry is a method for characterizing, counting, and separating cells that are suspended in a fluid. The basic flow cytometer uses hydrodynamic focusing to produce a very thin stream of fluid containing cells moving in single file through a quartz flow chamber (Fig. 20.2). The cells are charac-0 terized on the basis of their scattering and fluorescent properties. This simultaneous measurement of scattering and fluorescence is accomplished with a sophisticated optical system that detects light from the sample both at the wavelength of the excitation source (scattering) as well as at longer wavelengths (fluorescence) at more than one angle. Analysis of these measurements produces parameters related to the cells' size, granularity, and natural or tagged fluorescence. High-pressure mercury or xenon arc lamps can be used as light sources, but the argon laser (488 nm) is the preferred source for high-performance instruments.

0-8493-1813-0/03/$0.00+$ 1.50

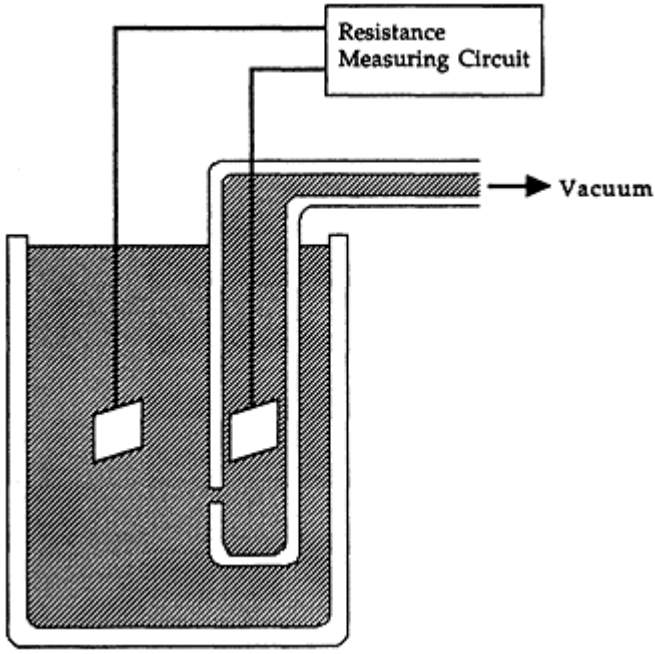


FIGURE 20.1 Coulter method. Blood cells are surrounded by an insulating membrane, which makes them non-conductive. The resistance of electrolyte-filled channel will increase slightly as cells flow through it. This resistance variation yields both the total number of cells that flow through the channel and the volume of each cell.

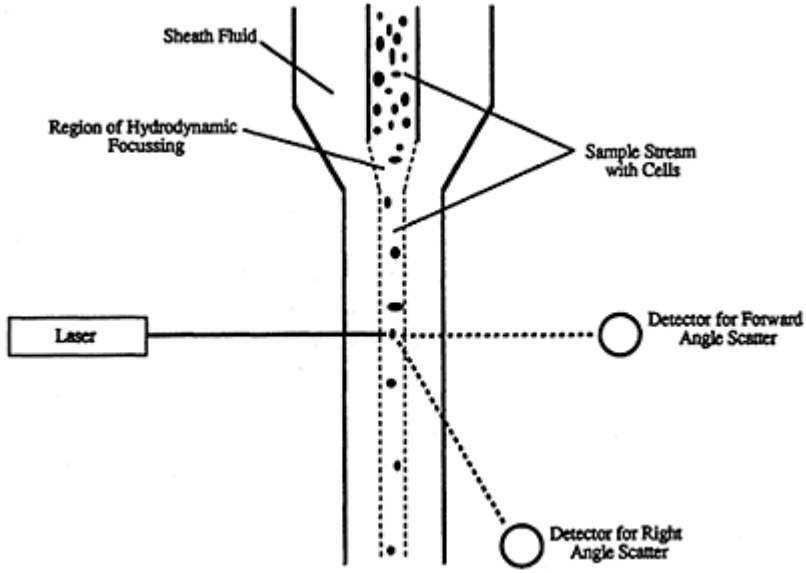


FIGURE 20.2 Flow cytometer. By combining hydrodynamic focusing, state-of-the-art optics, fluorescent labels, and high-speed computing, large numbers of cells can be characterized and sorted automatically.

One of the more interesting features of this technology is that particular cells may be selected at rates that allow collection of quantities of particular cell types adequate for further chemical testing. This is accomplished by breaking the outgoing stream into a series of tiny droplets using piezoelectric vibration. By charging the stream of droplets and then using deflection plates controlled by the cell analyzer, the cells of interest can be diverted into collection vessels.

The development of monoclonal antibodies coupled with flow cytometry allows for quantitation of T and B cells to assess the status of the immune system as well as characterization of leukemias, lymphomas, and other disorders.

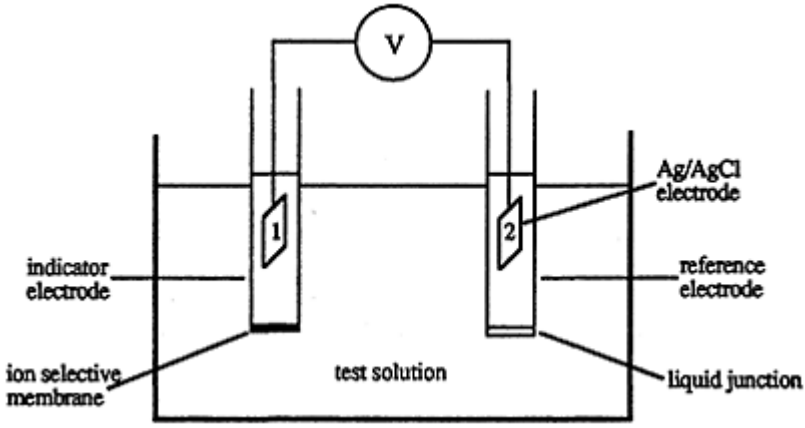


FIGURE 20.3 Electrochemical cell.

20.2 Electrochemical Methods

Electrochemical methods are increasingly popular in the clinical laboratory, for measurement not only of electrolytes, blood gases, and pH but also of simple compounds such as glucose. *Potentiometry* is a method in which a voltage is developed across electrochemical cells as shown in Fig. 20.3. This voltage is measured with little or no current flow.

Ideally, one would like to measure all potentials between the reference solution in the indicator electrode and the test solution. Unfortunately there is no way to do that. Interface potentials develop across any metal-liquid boundary, across liquid junctions, and across the ion-selective membrane. The key to making potentiometric measurements is to ensure that all the potentials are constant and do not vary with the composition of the test solution except for the potential of interest across the ion-selective membrane. By maintaining the solutions within the electrodes constant, the potential between these solutions and the metal electrodes immersed in them is constant. The liquid junction is a structure that severely limits bulk flow of the solution but allows free passage of all ions between the solutions. The reference electrode commonly is filled with saturated KCl, which produces a small, constant liquid-junction potential. Thus, any change in the measured voltage (V) is due to a change in the ion concentration in the test solution for which the membrane is selective.

The potential which develops across an ion-selective membrane is given by the Nernst equation:

$$V = \left(\frac{RT}{zF}\right) \ln \frac{a_2}{a_1}, \qquad (20.1)$$

where R=gas constant=8.314 J/K·mol

T=temperature in K

z=ionization number

F=Faraday constant=9.649×10^4 C/Mol

a_n=activity of ion in solution n.

When one of the solutions is a reference solution, this equation can be rewritten in a convenient form as

$$V = V_0 + \frac{N}{z}\log_{10} a\,, \tag{20.2}$$

where V_o=a constant voltage due to reference solution

N=Nernst slope ≈59 mV/decade at room temperature.

The actual Nernst slope is usually slightly less than the theoretical value. Thus, the typical pH meter has two calibration controls. One adjusts the offset to account for the value of V_o, and the other adjusts the range to account for both temperature effects and deviations from the theoretical Nernst slope.

20.3 Ion-Specific Electrodes

Ion-selective electrodes use membranes that are permeable only to the ion being measured. To the extent that this can be done, the specificity of the electrode can be very high. One way of overcoming a lack of specificity for certain electrodes is to make multiple simultaneous measurement of several ions that include the most important interfering ones. A simple algorithm can then make corrections for the interfering effects. This technique is used in some commercial electrolyte analyzers. A partial list of the ions that can be measured with ion-selective electrodes includes H^+ (pH), Na^+, K^+, Li^+, Ca^{++}, Cl^-, F^-, NH^+_4, and CO_2.

NH^+_4, and CO_2 are both measured with a modified ion-selective electrode. They use a pH electrode modified with a thin layer of a solution (sodium bicarbonate for CO_2 and ammonium chloride for NH^+_4) whose pH varies depending on the concentration of ammonium ions or CO_2 it is equilibrated with. A thin membrane holds the solution against the pH glass electrode and provides for equilibration with the sample solution. Note that the CO_2 electrode in Fig. 20.4 is a combination electrode. This means that both the reference and indicating electrodes have been combined into one unit. Most pH electrodes are made as combination electrodes.

The Clark electrode measures pO_2 by measuring the current developed by an electrode with an applied of approximately −0.65 V is applied to a platinum electrode relative to a Ag/AgCl electrode in an electrolyte voltage rather than a voltage measurement. This is an example of *amperometry*. In this electrode a voltage solution. The reaction,

$$O_2 + 2H^+ + 2e^- \rightarrow H_2O_2\,,$$

proceeds at a rate proportional to the partial pressure of oxygen in the solution. The electrons involved in this reaction form a current that is proportional to the rate of the reaction and thus to the pO_2 in the solution.

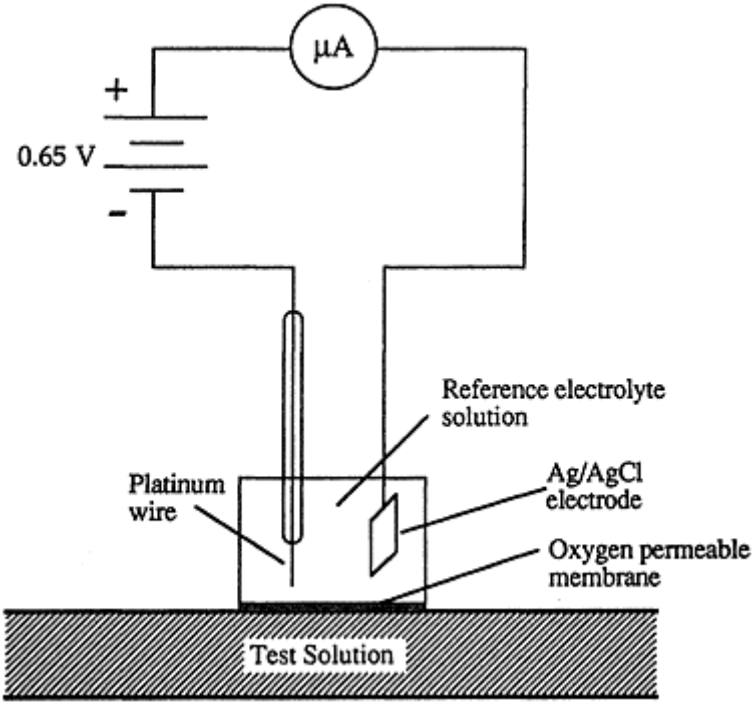


FIGURE 20.4 Clark electrode.

20.4 Radioactive Methods

Isotopes are atoms that have identical atomic number (number of protons) but different atomic mass numbers (protons+neutrons). Since they have the same number of electrons in the neutral atom, they have identical chemical properties. This provides an ideal method for labeling molecules in a way that allows for detection at extremely low concentrations. Labeling with radioactive isotopes is extensively used in radioimmunoassays where the amount of antigen bound to specific antibodies is measured. The details of radioactive decay are complex, but for our purposes there are three types of emission from decaying nuclei: *alpha, beta,* and *gamma radiation*. Alpha particles are made up of two neutrons and two protons (helium nucleus). Alpha emitters are rarely used in the clinical laboratory. Beta emission consists of electrons or positrons emitted from the nucleus. They have a continuous range of energies up to a maximum value characteristic of the isotope. Beta radiation is highly interactive with matter and cannot penetrate very far in most materials. Gamma radiation is a high-energy form of electromagnetic radiation. This type of radiation may be continuous, discrete, or mixed depending on the details of the decay process. It has greater penetrating ability than beta radiation. (See Fig. 20.5.)

The kinetic energy spectrum of emitted radiation is characteristic of the isotope. The energy is commonly measured in electron volts (eV). One electron volt is the energy acquired by an electron falling through a potential of 1 volt. The isotopes commonly used in the clinical laboratory have energy spectra that range from 18 keV−3.6 MeV.

The activity of a quantity of radioactive isotope is defined as the number of disintegrations that occur per second. The usual units are the curie (Ci), which is defined as 3.7×10^{10} dps, and the becquerel (Bq), defined as 1 dps. Specific activity for a given isotope is defined as activity per unit mass of the isotope.

The rate of decay for a given isotope is characterized by the decay constant λ, which is the proportion of the isotope which decays in unit time. Thus, the rate of loss of radioactive isotope is governed by the equation,

$$\frac{dN}{dt} = -\lambda N\,, \tag{20.3}$$

where N is the amount of radioactive isotope present at time t. The solution to this differential equation is:

$$N = N_0 e^{-\lambda t}\,. \tag{20.4}$$

It can easily be shown that the amount of radioactive isotope present will be reduced by half after time:

$$t_{1/2} = \frac{0.693}{\lambda}\,. \tag{20.5}$$

This is known as the half-life for the isotope and can vary widely; for example, carbon-14 has a half-life of 5760 years, and iodine-131 has a half-life of 8.1 days.

The most common method for detection of radiation in the clinical laboratory is by scintillation. This is the conversion of radiation energy into photons in the visible or near-UV range. These are detected with photomultiplier tubes.

For gamma radiation, the scintillating crystal is made of sodium iodide doped with about 1% thallium, producing 20 to 30 photons for each electron-volt of energy absorbed. The photomultiplier tube and amplifier circuit produce voltage pulses proportional to the energy of the absorbed radiation. These voltage pulses are usually passed through a pulse-height analyzer that eliminates pulses outside a preset energy range (window). Multichannel analyzers can discriminate between two or more isotopes if they have well-separated energy maxima. There generally will be some spill down of counts from the higher-energy isotope into the lower-energy isotope's window, but this effect can be corrected with a simple algorithm. Multiple well detectors with up to 64 detectors in an array are available that increase the throughput for counting systems greatly. Counters using the sodium iodide crystal scintillator are referred to as gamma counters or well counters.

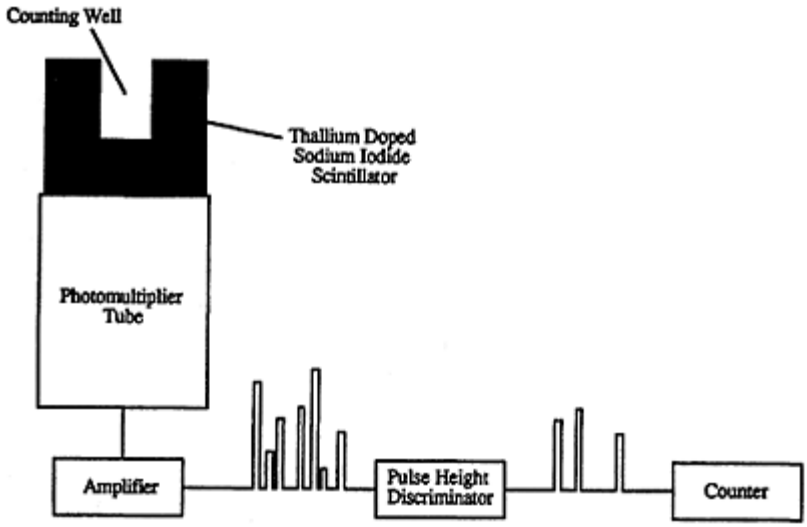


FIGURE 20.5 Gamma counter. The intensity of the light flash produced when a gamma photon interacts with a scintillator is proportional to the energy of the photon. The photomultiplier tube converts these light flashes into electric pulses that can be selected according to size (gamma energy) and counted.

The lower energy and short penetration ability of beta particles requires a scintillator in direct contact with the decaying isotope. This is accomplished by dissolving or suspending the sample in a liquid fluor. Counters that use this technique are called beta counters or liquid scintillation counters.

Liquid scintillation counters use two photomultiplier tubes with a coincidence circuit that prevents counting of events seen by only one of the tubes. In this way, false counts due to chemiluminescence and noise in the phototube are greatly reduced. Quenching is a problem in all liquid scintillation counters. Quenching is any process that reduces the efficiency of the scintillation counting process, where efficiency is defined as

Efficiency=counts per minute/decays per minute.

(20.6)

A number of techniques have been developed that automatically correct for quenching effects to produce estimates of true decays per minute from the raw counts. Currently there is a trend away from beta-emitting isotopic labels, but these assays are still used in many laboratories.

20.5 Coagulation Timers

Screening for and diagnosis of coagulation disorders is accomplished by assays that determine how long it takes for blood to clot following initiation of the clotting cascade by various reagents. A variety of instruments have been designed to automate this procedure. In addition to increasing the speed and throughput of such testing, these instruments improve the reproducibility of such tests. All the instruments provide precise introduction of reagents, accurate timing circuits, and temperature control. They differ in the method for detecting clot formation. One of the older methods still in use is to dip a small metal hook into the blood sample repeatedly and lift it a few millimeters above the surface. The electric resistance between the hook and the sample is measured, and when fibrin filaments form, they produce a conductive pathway that is detected as clot formation. Other systems detect the increase in viscosity due to fibrin formation or the scattering due to the large polymerized molecules formed. Absorption and fluorescence spectroscopy can also be used for clot detection.

20.6 Osmometers

The *colligative properties* of a solution are a function of the number of solute particles present regardless of size or identity. Increased solute concentration causes an increase in osmotic pressure and boiling point and a decrease in vapor pressure and freezing-point. Measuring these changes provides information on the total solute concentration regardless of type. The most accurate and popular method used in clinical laboratories is the measurement of freezing point depression. With this method, the sample is supercooled to a few degrees below 0°C while being stirred gently. Freezing is then initiated by vigorous stirring. The heat of fusion quickly brings the solution to a slushy state where an equilibrium exists between ice and liquid, ensuring that the temperature is at the freezing-point. This temperature is measured. A solute concentration of 1 osmol/kg water produces a freezing point depression of 1.858°C. The measured temperature depression is easily calibrated in units of milliosmols/kg water.

The vapor-pressure depression method has the advantage of smaller sample size. However, it is not as precise as the freezing-point method and cannot measure the contribution of volatile solutes such as ethanol. This method is not used as widely as the freezing-point depression method in clinical laboratories.

Osmolality of blood is primarily due to electrolytes such as Na^+ and Cl^-. Proteins with molecular weights of 30,000 or more atomic mass units (amu) contribute very little to total osmolality due to their smaller numbers (a single Na^+ ion contributes just as much to osmotic pressure as a large protein molecule). However, the contribution to osmolality made by proteins is of great interest when monitoring conditions leading to pulmonary edema. This value is known as colloid osmotic pressure, or oncotic pressure, and is measured with a membrane permeable to water and all molecules smaller than about 30,000 amu. By placing a reference saline solution on one side and the unknown sample on the other, an osmotic pressure is developed across the membrane. This pressure is measured with a pressure transducer and can be related to the true colloid osmotic pressure through a calibration procedure using known standards.

20.7 Automation

Improvements in technology coupled with increased demand for laboratory tests as well as pressures to reduce costs have led to the rapid development of highly automated laboratory instruments. Typical automated instruments contain mechanisms for measuring, mixing, and transport of samples and reagents, measurement systems, and one or more microprocessors to control the entire system. In addition to system control, the computer systems store calibration curves, match test results to specimen IDs, and generate reports. Automated instruments are dedicated to complete blood counts, coagulation studies, microbiology assays, and immunochemistry, as well as high-volume instruments used in clinical chemistry laboratories. The chemistry analyzers tend to fall into one of four classes: continuous flow, centrifugal, pack-based, and dry-slide-based systems. The continuous flow systems pass successive samples and reagents through a single set of tubing, where they are directed to appropriate mixing, dialyzing, and measuring stations. Carryover from one sample to the next is minimized by the introduction of air bubbles and wash solution between samples.

Centrifugal analyzers use plastic rotors that serve as reservoirs for samples and reagents and also as cuvettes for optical measurements. Spinning the plastic rotor mixes, incubates, and transports the test solution into the cuvette portion of the rotor, where the optical measurements are made while the rotor is spinning.

Pack-based systems are those in which each test uses a special pack with the proper reagents and sample preservation devices builtin. The sample is automatically introduced into as many packs as tests required. The packs are then processed sequentially.

Dry chemistry analyzers use no liquid reagents. The reagents and other sample preparation methods are layered onto a slide. The liquid sample is placed on the slide, and after a period of time the color developed is read by reflectance photometry. Ion-selective electrodes have been incorporated into the same slide format.

There are a number of technological innovations found in many of the automated instruments. One innovation is the use of fiberoptic bundles to channel excitation energy toward the sample as well as transmitted, reflected, or emitted light away from the sample to the detectors. This provides a great deal of flexibility in instrument layout. Multiwavelength analysis using a spinning filter wheel or diode array detectors is commonly found. The computers associated with these instruments allow for innovative improvements in the assays. For instance, when many analytes are being analyzed from one sample, the interference effects of one analyte on the measurement of another can be predicted and corrected before the final report is printed.

20.8 Trends in Laboratory Instrumentation

Predicting the future direction of laboratory instrumentation is difficult, but there seem to be some clear trends. Decentralization of the laboratory functions will continue with more instruments being located in or around ICUs, operating rooms, emergency rooms, and physician offices. More electrochemistry-based tests will be developed. The flame photometer is already being replaced with ion-selective electrode methods. Instruments that analyze whole blood rather than *plasma* or *serum* will reduce the amount of time

required for sample preparation and will further encourage testing away from the central laboratory. Dry reagent methods increasingly will replace wet chemistry methods. Radioimmunoassays will continue to decline with the increasing use of methods for performing immunoassays that do not rely upon radioisotopes such as enzyme-linked fluorescent assays.

Defining Terms

Alpha radiation: Particulate radiation consisting of a helium nucleus emitted from a decaying anucleus.
Amperometry: Measurements based on current flow produced in an electrochemical cell by an applied voltage.
Beta radiation: Particulate radiation consisting of an electron or positron emitted from a decaying nucleus.
Colligative properties: Physical properties that depend on the number of molecules present rather than on their individual properties.
Gamma radiation: Electromagnetic radiation emitted from an atom undergoing nuclear decay.
Hydrodynamic focusing: A process in which a fluid stream is first surrounded by a second fluid and then narrowed to a thin stream by a narrowing of the channel.
Isotopes: Atoms with the same number of protons but differing numbers of neutrons.
Plasma: The liquid portion of blood.
Potentiometry: Measurement of the potential produced by electrochemical cells under equilibrium conditions with no current flow.
Scintillation: The conversion of the kinetic energy of a charged particle or photon to a flash of light.
Serum: The liquid portion of blood remaining after clotting has occurred.

References

Burtis CA, Ashwood ER (eds). 1994. *Tietz Textbook of Clinical Chemistry*, 2d ed, Philadelphia, Saunders Company.
Hicks MR, Haven MC, Schenken JR, et al. (eds). 1987. *Laboratory Instrumentation*, 3d ed, Philadelphia, Lippincott Company, 1987.
Kaplan LA, Pesce AJ (eds). 1989. *Clinical Chemistry: Theory, Analysis, and Correlation*, 2d ed, St. Louis, Mosby.
Tietz NW (ed). 1987. *Fundamentals of Clinical Chemistry*, 3d ed, Philadelphia, Saunders.
Ward JM, Lehmann CA, Leiken AM. 1994. *Clinical Laboratory Instrumentation and Automation: Principles, Applications, and Selection*, Philadelphia, Saunders.

21
Implantable Cardiac Pacemakers

Michael Forde
Medtronic, inc.

Pat Ridgely
Medtronic, Inc.

The practical use of an implantable device for delivering a controlled, rhythmic electric stimulus to maintain the heartbeat is relatively recent: Cardiac pacemakers have been in clinical use only slightly more than 30 years. Although devices have gotten steadily smaller over this period (from 250 grams in 1960 to 25 grams today), the technological evolution goes far beyond size alone. Early devices provided only single-chamber, asynchronous, *nonprogrammable* pacing coupled with questionable reliability and longevity. Today, advanced electronics afford dual-chamber *multiprogrammability*, diagnostic functions, rate response, data collection, and exceptional reliability, and lithium-iodine power sources extend longevity to upward of 10 years. Continual advances in a number of clinical, scientific, and engineering disciplines have so expanded the use of pacing that it now provides cost-effective benefits to an estimated 350,000 patients worldwide each year.

The modern pacing system is comprised of three distinct components: pulse generator, lead, and programmer (Fig. 21.1). The pulse generator houses the battery and the circuitry that generates the stimulus and senses electrical activity. The lead is an insulated wire that carries the stimulus from the generator to the heart and relays intrinsic cardiac signals back to the generator. The programmer is a telemetry device used to provide two-way communications between the generator and the clinician. It can alter the therapy delivered by the pacemaker and retrieve diagnostic data that are essential for optimally titrating that therapy. Ultimately, the therapeutic success of the pacing prescription rests on the clinician's choice of an appropriate system, use of sound implant technique, and programming focused on patient outcomes.

This chapter discusses in further detail the components of the modern pacing system and the significant evolution that has occurred since its inception. Our focus is on system design and operations, but we also briefly overview issues critical to successful clinical performance.

0-8493-1813-0/03/$0.00+$ 1.50

21.1 Indications

The decision to implant a permanent pacemaker for bradyarrhythmias usually is based on the major goals of symptom relief (at rest and with physical activity), restoration of functional capacity and quality

FIGURE 21.1 The pacing systems comprise a programmer, pulse generator, and lead. Two programmers are pictured above; one is portable, and the other is an office-based unit.

of life, and reduced mortality. As with other healthcare technologies, appropriate use of pacing is the intent of indications guidelines established by Medicare and other third-party payors.

In 1984 and again in 1991, a joint commission of the American College of Cardiology and the American Heart Association established guidelines for pacemaker implantation [Committee on Pacemaker Implantation, 1991]. In general, pacing is indicated when there is a dramatic slowing of the heart rate or a failure in the connection between the atria and ventricles resulting in decreased cardiac output manifested by such symptoms as syncope, lightheadedness, fatigue, and exercise intolerance. Failure of impulse formation and/or conduction is the overriding theme of all pacemaker indications. There are four categories of pacing indications:

1. Heart block (e.g., complete heart block, symptomatic 2° AV block)

2. Sick sinus syndrome (e.g., symptomatic bradycardia, sinus arrest, sinus exit block)
3. Myocardial infarction (e.g., conduction disturbance related to the site of infarction)
4. Hypersensitive carotid sinus syndrome (e.g., recurrent syncope)

Within each of these four categories the ACC/AHA provided criteria for classifying a condition as group I (pacing is considered necessary), group II (pacing may be necessary), or group III (pacing is considered inappropriate).

New indications for cardiac pacing are being evaluated under the jurisdiction of the Food and Drug Administration. For example, *hypertrophic obstructive cardiomyopathy* (HOCM) is one of these new potential indications, with researchers looking at dual-chamber pacing as a means of reducing left ventricular outflow obstruction. Though efforts in these areas are ongoing and expanding, for now they remain unapproved as standard indications for pacing.

21.2 Pulse Generators

The pulse generator contains a power source, output circuit, sensing circuit, and a timing circuit (Fig. 21.2). A telemetry coil is used to send and receive information between the generator and the programmer. *Rate-adaptive* pulse generators include the sensor components along with the circuit to process the information measured by the sensor.

Modern pacemakers use *CMOS circuit* technology. One to 2 kilobytes of read-only memory (ROM) are used to direct the output and sensing circuits; 16 to 512 bytes of random-access memory (RAM) are

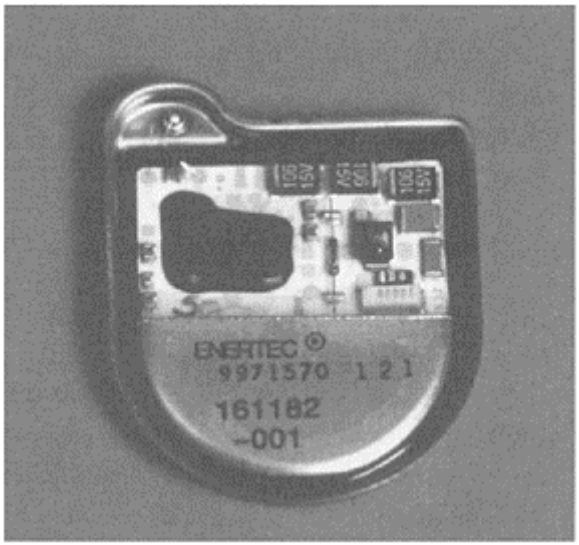


FIGURE 21.2 Internal view of pulse generator.

used to store diagnostic data. Some manufacturers offer fully RAM-based pulse generators, providing greater storage of diagnostic data and the flexibility for changing feature sets after implantation.

All components of the pulse generator are housed in a *hermetically* sealed titanium case with a connector block that accepts the lead(s). Because pacing leads are available with a variety of different connector sites and configurations, the pulse generator is available with an equal variety of connectors. The outer casing is laser-etched with the manufacturer, name, type (e.g., single- versus dual-chamber), model number, serial number, and the lead connection diagram for each identification. Once implanted, it may be necessary to use an x-ray to reveal the identity of the generator. Some manufacturers use radiopaque symbols and ID codes for this purpose, whereas others give their generators characteristic shapes.

Sensing Circuit

Pulse generators have two basic functions, pacing and sensing. Sensing refers to the recognition of an appropriate signal by the pulse generator. This signal is the intrinsic cardiac depolarization from the chamber or chambers in which the leads are placed. It is imperative for the sensing circuit to discriminate between these intracardiac signals and unwanted electrical interference such as far-field cardiac events, diastolic potentials, skeletal muscle contraction, and pacing stimuli. An intracardiac electrogram (Fig. 21.3) shows the waveform as seen by the pacemaker; it is typically quite different from the corresponding event as shown on the surface EGG.

Sensing (and pacing) is accomplished with one of two configurations, bipolar and unipolar. In bipolar, the anode and cathode are close together, with the anode at the tip of the lead and the cathode a ring electrode about 2 cm proximal to the tip. In unipolar, the anode and cathode may be 5 to 10 cm apart. The anode is at the lead tip and the cathode is the pulse generator itself (usually located in the pectoral region).

In general, bipolar and unipolar sensing configurations have equal performance. A drawback of the unipolar approach is the increased possibility of sensing noncardiac signals: The large electrode separation may, for example, sense myopotentials from skeletal muscle movement, leading to inappropriate inhibition of pacing. Many newer pacemakers can be programmed to sense or pace in either configuration.

Once the electrogram enters the sensing circuit, it is scrutinized by a bandpass filter (Fig. 21.4). The frequency of an R-wave is 10 to 30 Hz. The center frequency of most sensing amplifiers is 30 Hz. T-waves are slower, broad signals that are composed of lower frequencies (approximately 5 Hz or less). Far-field signals are also lower-frequency signals, whereas skeletal muscle falls in the range of 10 to 200 Hz.

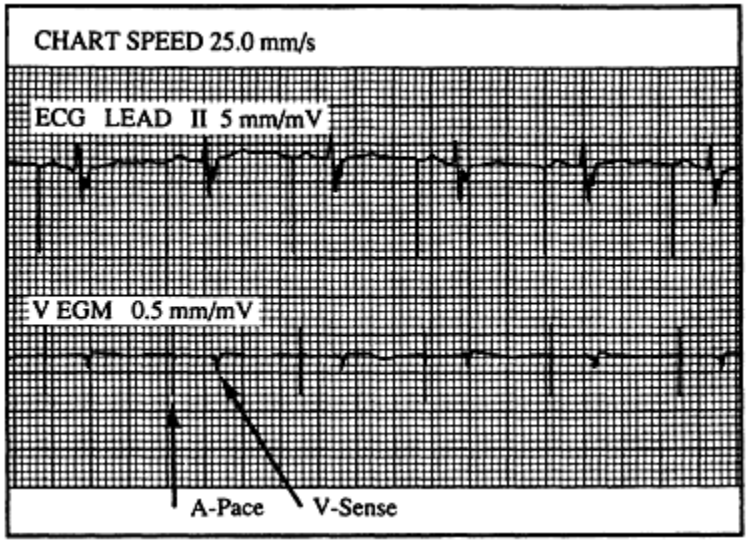


FIGURE 21.3 The surface EGG (EGG LEAD II) represents the sum total of the electrical potentials of all depolarizing tissue. The intracardiac electrogram (VEGM) shows only the potentials measured between the lead electrodes. This allows the evaluation of signals that may be hidden within the surface EGG.

At the implant, the voltage amplitude of the *R*-wave (and the *P*-wave, in the case of dual-chamber pacing) is measured to ensure the availability on an adequate signal. *R*-wave amplitudes are typically 5 to 25 mV, and *P*-wave amplitudes are 2 to 6 mV. The signals passing through the sense amplifier are compared to an adjustable reference voltage called the *sensitivity*. Any signal below the reference voltage is not sensed, and those above it are sensed. Higher-sensitivity settings (high-reference voltage) may lead to substandard sensing, and a lower reference voltage may result in oversensing. A minimum 2:1 safety margin should be maintained between the sensitivity setting and the amplitude of the intracardiac signal. The circuit is protected from extremely high voltages by a Zener diode.

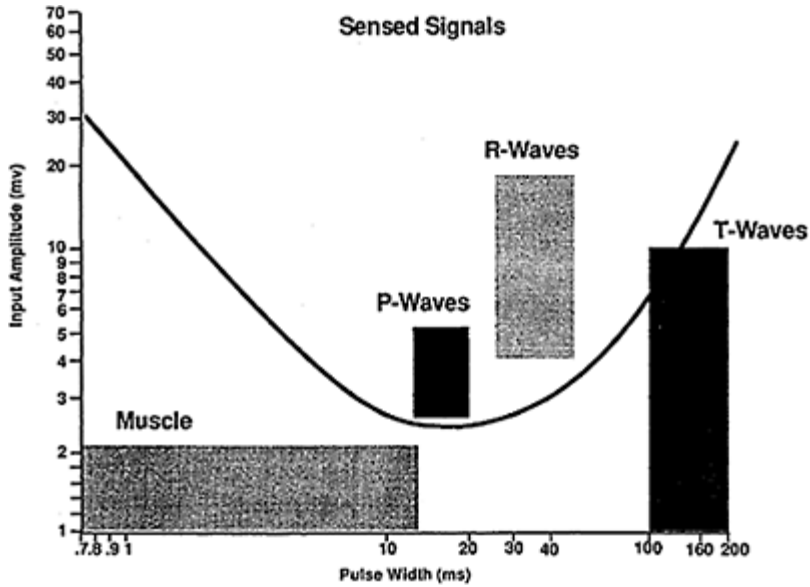


FIGURE 21.4 This is a conceptual depiction of the bandpass filter demonstrating the typical filtering of unwanted signals by discriminating between those with slew rates that are too low and/or too high.

The slope of the signal is also surveyed by the sensing circuit and is determined by the slew rate (the time rate of change in voltage). A slew rate that is too flat or too steep may be eliminated by the bandpass filter. On the average, the slew rate measured at implant should be between 0.75 and 2.50 V/s.

The last line of defense in an effort to remove undesirable signals is to "blind" the circuit at specific times during the cardiac cycle. This is accomplished with blanking and refractory periods. Some of these periods are programmable. During the blanking period the sensing circuit is turned off, and during the refractory period the circuit can see the signal but does not initiate any of the basic timing intervals. Virtually all paced and sensed events begin concurrent blanking and refractory periods, typically ranging from 10 to 400 ms. These are especially helpful in dual-chamber pacemakers where there exists the potential for the pacing output of the atrial side to inhibit the ventricular pacing output, with dangerous consequences for patients in complete heart block.

Probably the most common question asked by the general public about pacing systems is the effect of electromagnetic interference (EMI) on their operation. EMI outside of the hospital is an infrequent problem, though patients are advised to avoid such sources of strong electromagnetic fields as arc welders, high-voltage generators, and radar antennae. Some clinicians suggest that patients avoid standing near antitheft devices used in retail

stores. Airport screening devices are generally safe, though they may detect a pacemaker's metal case. Microwave ovens, ham radio equipment, video games, computers, and office equipment rarely interfere with the operation of modern pacemakers. A number of medical devices and procedures may on occasion do so, however; electrocautery, cardioversion and defibrillation, MRI, lithotripsy, diathermy, TENS units, and radiation therapy.

Pacemakers affected by interference typically respond with temporary loss of output or temporary reversion to asynchronous pacing (pacing at a fixed rate, with no inhibition from intrinsic cardiac events). The usual consequence for the patient is a return of the symptoms that originally led to the pacemaker implant.

Output Circuit

Pacing is the most significant drain on the pulse generator power source. Therefore, current drain must be minimized while maintaining an adequate safety margin between the *stimulation threshold* and the programmed output stimulus. Modern permanent pulse generators use constant voltage. The voltage remains at the programmed value while current fluctuates in relation to the source impedance.

Output energy is controlled by two programmable parameters, pulse amplitude and pulse duration. Pulse amplitudes range from 0.8 to 5 V and, in some generators, can be as high as 10 V (used for troubleshooting or for pediatric patients). Pulse duration can range from 0.05 to 1.5 ms. The prudent selection of these parameters will greatly influence the longevity of the pulse generator.

The output pulse is generated from the discharge of a capacitor charged by the battery. Most modern pulse generators contain a 2.8-V battery. The higher voltages are achieved using voltage multipliers (smaller capacitors used to charge the large capacitor). The voltage can be doubled by charging two smaller capacitors in parallel, with the discharge delivered to the output capacitor in series. Output pulses are emitted at a rate controlled by the timing circuit; output is commonly inhibited by sensed cardiac signals.

Timing Circuit

The timing circuit regulates such parameters as the pacing cycle length, refractory and blanking periods, pulse duration, and specific timing intervals between atrial and ventricular events. A crystal oscillator generating frequencies in the kHz range sends a signal to a digital timing and logic control circuit, which in turn operates internally generated clocks at divisions of the oscillatory frequency.

A rate-limiting circuit is incorporated into the timing circuit to prevent the pacing rate from exceeding an upper limit should a random component failure occur (an extremely rare event). This is also referred to as "runaway" protection and is typically 180 to 200 ppm.

Telemetry Circuit

Today's pulse generators are capable of both transmitting information from an RF antenna and receiving information with an RF decoder. This two-way communication

occurs between the pulse generator and the programmer at approximately 300 Hz. Real-time telemetry is the term used to describe the ability of the pulse generator to provide information such as pulse amplitude, pulse duration, lead impedance, battery impedance, lead current, charge, and energy. The programmer, in turn, delivers coded messages to the pulse generator to alter any of the programmable features and to retrieve diagnostic data. Coding requirements reduce the likelihood of inappropriate programming alterations by environmental sources of radiofrequency and magnetic fields. It is also prevents the improper use of programmers from other manufacturers.

Power Source

Over the years, a number of different battery technologies have been tried, including mercury-zinc, rechargeable silver-modified-mercuric-oxide-zinc, rechargeable nickel-cadmium, radioactive plutonium or promethium, and lithium with a variety of different cathodes. Lithium-cupric-sulfide and mercury-zinc batteries were associated with corrosion and early failure. Mercury-zinc produced hydrogen gas as a by-product of the battery reaction; the venting required made it impossible to hermetically seal the generator. This led to fluid infiltration followed by the risk of sudden failure.

The longevity of very early pulse generators was measured in hours. With the lithium-iodide technology now used, longevity has been reported as high as 15 years. The clinical desire to have a generator that is small and full-featured yet also long-lasting poses a formidable challenge to battery designers. One response by manufacturers has been to offer different models of generators, each offering a different balance between therapy, size, and longevity. Typical *battery capacity* is in the range of 0.8 to 3.0 amp-hours.

Many factors affect longevity, including pulse amplitude and duration, pacing rate, single- versus dual-chamber pacing, degree to which the patient uses the pacemaker, lead design, and static current drain from the sensing circuits. Improvements in lead design are often overlooked as a factor in improving longevity, but electrodes used in 1960 required a pulse generator output of 675 μJ for effective stimulation, whereas the electrodes of the 1990s need only 3 to 6 μJ.

Another important factor in battery design lies in the electrolyte that separates the anode and the cathode. The semisolid layer of lithium iodide that is used gradually thickens over the life of the cell, increasing the internal resistance of the battery. The voltage produced by lithium-iodine batteries is inversely related to this resistance and is linear from 2.8 V to approximately 2.4 V, representing about 90% of the usable battery life. It then declines exponentially to 1.8 V as the internal battery resistance increases from 10,000 Ω to 40,000 Ω (Fig. 21.5).

When the battery reaches between 2.0 and 2.4 V (depending on the manufacturer), certain functions of the pulse generator are altered so as to alert the clinician. These alterations are called the electivereplacement indicators (ERI). They vary from one pulse generator to another and include signature decreases in rate, a change to a specific pacing *mode*, pulse duration stretching, and the telemetered battery voltage. When the battery voltage reaches 1.8 V, the pulse generator may operate erratically or cease to function and is said to have reached “end of life.” The time period between appearance of the ERI and end-of-life status averages about 3 to 4 months.

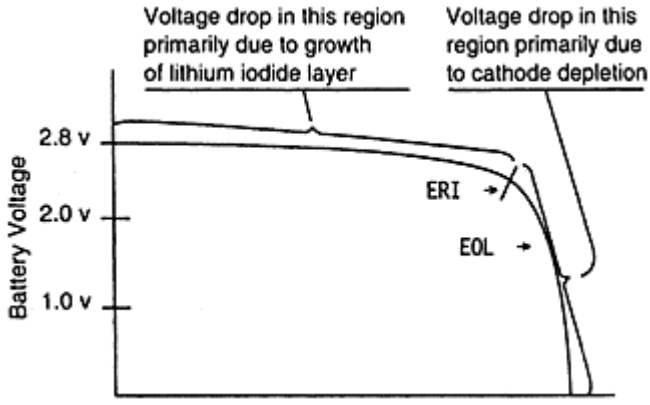


FIGURE 21.5 The initial decline in battery voltage is slow and then more rapid after the battery reaches the ERI voltage. An important aspect of battery design is the predictability of this decline so that timely generator replacement is anticipated.

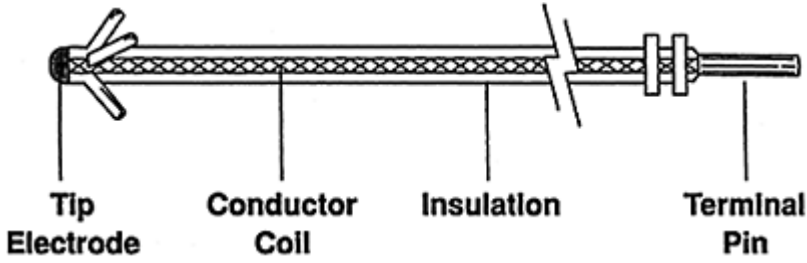


FIGURE 21.6 The four major lead components.

21.3 Leads

Implantable pacing leads must be designed not only for consistent performance within the hostile environment of the body but also for easy handling by the implanting physician. Every lead has four major components (Fig. 21.6): the electrode, the conductor, the insulation, and the connector pin(s).

The electrode is located at the tip of the lead and is in direct contact with the myocardium. Bipolar leads have a tip electrode and a ring electrode (located about 2 cm proximal to the tip); unipolar leads have tip electrodes only. A small-radius electrode

provides increased current density resulting in lower stimulation thresholds. The electrode also increases resistance at the electrode-myocardial interface, thus lowering the current drain further and improving battery longevity. The radius of most electrodes is 6 to 8 mm^2, though there are clinical trials underway using a "high-impedance" lead with a tip radius as low as 1.5 mm^2.

Small electrodes, however, historically have been associated with inferior sensing performance. Lead designers were able to achieve both good pacing and good sensing by creating porous-tip electrodes containing thousands of pores in the 20–100 (im range. The pores allow the ingrowth of tissue, resulting in the necessary increase in effective sensing area while maintaining a small pacing area. Some commonly used electrode materials include platinum-iridium. Elgiloy (an alloy of cobalt, iron, chromium, molybdenum, nickel, and manganese), platinum coated with platinized titanium, and vitreous or pyrolytic carbon coating a titanium or graphite core.

Another major breakthrough in lead design is the steroid-eluting electrode. About 1 mg of a corticosteroid (dexamethasone sodium phosphate) is contained in a silicone core that is surrounded by the electrode material (Fig. 21.7). The "leaking" of the steroid into the myocardium occurs slowly over several years and reduces the inflammation that results from the lead placement. It also retards the growth of the fibrous sack that forms around the electrode that separates it from viable myocardium. As a result, the dramatic rise in acute thresholds that is seen with nonsteroid leads over the 8 to 16 weeks postimplant is nearly eliminated. This makes it possible to program a lower pacing output, further extending longevity.

Once a lead has been implanted, it must remain stable (or fixated). The fixation device is either active or passive. The active fixation leads incorporate corkscrew mechanisms, barbs, or hooks to attach themselves to the myocardium. The passive fixation leads are held into place with tines that become entangled into the netlike lining (trabeculae) of the heart. Passive leads generally have better acute pacing and sensing performance but are difficult to remove chronically. Active leads are easier to remove chronically

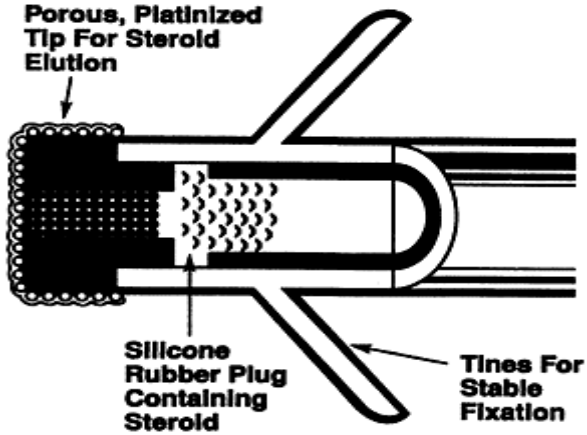


FIGURE 21.7 The steroid elution electrode.

and have the advantage of unlimited placement sites. Some implanters prefer to use active-fixation leads in the atrium and passive-fixation leads in the ventricle.

The conductor carries electric signals to the pulse generator and delivers the pacing pulses to the heart. It must be strong and flexible to withstand the repeated flexing stress placed on it by the beating heart. The early conductors were a single, straight wire that was vulnerable to fracturing. They have evolved into coiled (for increased flexibility) multifilar (to prevent complete failure with partial fractures) conductors. The conductor material is a nickel alloy called MP35N. Because of the need for two conductors, bipolar leads are usually larger in diameter than unipolar leads. Current bipolar leads have a coaxial design that has significantly reduced the diameter of bipolar leads.

Insulation materials (typically silicone and polyurethane) are used to isolate the conductor. Silicone has a longer history and the exclusive advantage of being repairable. Because of low tear strength, however, silicone leads tend to be thicker than polyurethane leads. Another relative disadvantage of silicone is its high coefficient of friction in blood, which makes it difficult for two leads to pass through the same vein. A coating applied to silicone leads during manufacturing has diminished this problem.

A variety of generator-lead connector configurations and adapters are available. Because incompatibility can result in disturbed (or even lost) pacing and sensing, an international standards (IS-1) has been developed in an attempt to minimize incompatibility.

Leads can be implanted epicardially and endocardially. *Epicardial* leads are placed on the outer surface of the heart and require the surgical exposure of a small portion of the heart. They are used when venous occlusion makes it impossible to pass a lead transvenously, when abdominal placement of the pulse generator is needed (as in the case of radiation therapy to the pectoral area), or in children (to allow for growth). *Endocardial* leads are more common and perform better in the long term. These leads are passed through the venous system and into the right side of the heart. The subclavian or cephalic veins in the pectoral region are common entry sites. Positioning is facilitated by a thin, firm wire stylet that passes through the central lumen of the lead, stiffening it. Fluoroscopy is used to visualize lead positioning and to confirm the desired location.

Manufacturers are very sensitive to the performance reliability of the leads. Steady improvements in materials, design, manufacturing, and implant technique have led to reliability well in excess of 99% over 3-year periods.

21.4 Programmers

Noninvasive reversible alteration of the functional parameters of the pacemaker is critical to ongoing clinical management. For a pacing system to remain effective throughout its lifetime, it must be able to adjust to the patient’s changing needs. The programmer is the primary clinical tool for changing settings, for retrieving diagnostic data, and for conducting noninvasive tests.

The pacing rate for programmable pacemakers of the early 1960s was adjusted via a Keith needle manipulated percutaneously into a knob on the side of the pacemaker; rotating the needle changed the pacing rate. Through the late 1960s and early 1970s, magnetically attuned reed switches in the pulse generator made it possible to

noninvasively change certain parameters such as rate, output, sensitivity, and polarity. The application of a magnet could alter the parameters that were usually limited to only one of two choices. It wasn't until the late 1970s, when radiofrequency energy was incorporated as the transmitter of information, that programmability began to realize its full potential. Radiofrequency transmission is faster, provides bidirectional telemetry, and decreases the possibility of unintended programming from inappropriate sources.

Most manufacturers today are moving away from a dedicated proprietary instrument and toward a PC-based design. The newer designs are generally more flexible, more intuitive to use, and more easily updated when new devices are released. Manufacturers and clinicians alike are becoming more sensitive to the role that time-efficient programming can play in the productivity of pacing clinics, which may provide follow-up for as many as 500 to 1000 patients a year.

21.5 System Operation

Pacemakers have gotten steadily more powerful over the last three decades, but at the cost of steadily greater complexity. Manufacturers have come to realize the challenge that this poses for busy clinicians and have responded with a variety of interpretive aids (Fig. 21.8).

Much of the apparent complexity of the timing rules that determine pacemaker operation is due to a design goal of mimicking normal cardiac function without interfering with it. One example is the dual-chamber feature that provides sequential stimulation of the atrium before the ventricle.

Another example is rate response, designed for patients who lack the normal ability to increase their heart rate in response to a variety of physical conditions (e.g., exercise). Introduced in the mid-1980s, rate-responsive systems use some sort of sensor to measure the change in a physical variable correlated to heart rate. The sensor output is signal-processed and then used by the output circuit to specify a target pacing rate. The clinician controls the aggressiveness of the rate increase through a variety of parameters (including a choice of transfer function); pacemaker-resident diagnostics provide data helpful in titrating the rate-response therapy.

The most common sensor is the activity sensor, which uses piezoelectric materials to detect vibrations caused by body movement. Systems using a transthoracic-impedance sensor to estimate pulmonary *minute ventilation* are also commercially available. Numerous other sensors (e.g., stroke volume, blood temperature or pH, oxygen saturation, pre-ejection interval, right ventricular pressure) are in various stages of clinical research or have been market released outside the United States. Some of these systems are dual-sensor, combining the best features of each sensor in a single pacing system.

To make it easier to understand the gross-level system operation of modern pacemakers, a five-letter code has been developed by the North American Society of Pacing and Electrophysiology and the British Pacing and Electrophysiology Group [Bernstein et al, 1987]. The first letter indicates the chamber (or chambers) that are paced. The second letter reveals those chambers in which sensing takes place, and the third letter describes how the pacemaker will respond to a sensed event. The pacemaker

will “inhibit” the pacing output when intrinsic activity is sensed or will “trigger” a pacing output based on a specific previously sensed event. For example, in DDD mode:

D: Pacing takes place in the atrium and the ventricle.

D: Sensing takes place in the atrium and the ventricle.

D: Both inhibition and triggering are the response to a sensed event. An atrial output is inhibited with an atrial-sensed event, whereas a ventricular output is inhibited with a ventricular-sensed event; a ventricular pacing output is triggered by an atrial-sensed event (assuming no ventricular event occurs during the A-V interval).

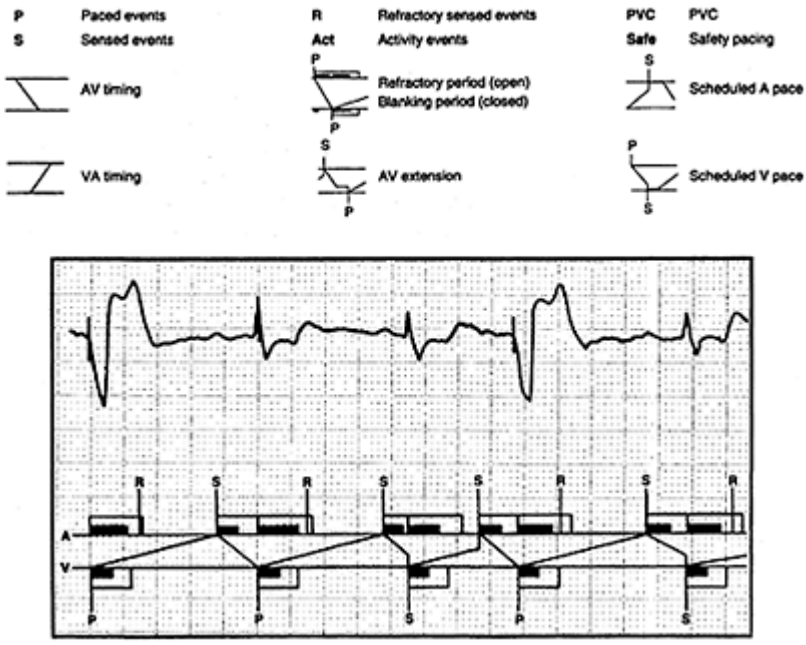


FIGURE 21.8 The Marker Channel Diagram is just one tool that makes interpretation of the ECG strip faster and more reliable for the clinician. It allows quick checking of the timing operations of the system.

The fourth letter in the code is intended to reflect the degree of programmability of the pacemaker but is typically used to indicate that the device can provide rate response. For example, a DDDR device is one that is programmed to pace and sense in both chambers and is capable of sensor-driven rate variability. The fifth letter is reserved specifically for antitachycardia functions (Table 21.1).

Readers interested in the intriguing details of pacemaker timing operations are referred to the works listed at the end of this chapter.

21.6 Clinical Outcomes and Cost Implications

The demonstrable hemodynamic and symptomatic benefits provided by rate-responsive and dual-chamber pacing have led U.S. physicians to include at least one of these features in over three fourths of implants in

TABLE 21.1 The NASPE/NPEG Code

Position	I	II	III	IV	V
Category	Chamber(s) paced	Chamber(s) sensed	Response to sensing	Programmability rate modulation	Antitachyarrhythmia function(s)
	O=None	O=None	O=None	O=None	O=None
	A=Atrium	A=Atrium	T=Triggered	P=Simple programmable	P=Packing
	V=Ventricle	V=Ventricle	I=Inhibited	M=Multiprogrammable	S=Shock
	D=Dual (A+V)	D=Dual (A+V)	D=Dual (T+I)	C=Communicating R=Rate modulation	D=Dual (P+S)
Manufacturers' S=Single (A or V) S=Single (A or V) designation only					

Note: Positions I through III are used exclusively for antibradyarrhythmia function. (From Bernstein AD, et al., *PACE,* Vol. 10, July–Aug. 1987.)

recent years. Also, new prospective data [Andersen et al., 1993] support a hypothesis investigated retrospectively since the mid-1980s: namely, that pacing the atrium in patients with sinus node dysfunction can dramatically reduce the incidence of such life-threatening complications as *congestive heart failure* and stroke associated with chronic *atrial fibrillation*. Preliminary analysis of the cost implications suggest that dualchamber pacing is significantly cheaper to the U.S. healthcare system than is single-chamber pacing over the full course of therapy, despite the somewhat higher initial cost of implanting the dual-chamber system.

21.7 Conclusion

Permanent cardiac pacing is the beneficiary of three decades of advances in a variety of key technologies: biomaterials, electrical stimulation, sensing of bioelectrical events, power sources, microelectronics, transducers, signal analysis, and software development. These advances, informed and guided by a wealth of clinical experience acquired during that time, have made pacing a cost-effective cornerstone of cardiac arrhythmia management.

Defining Terms

Atrial fibrillation: An atrial arrhythmia resulting in chaotic current flow within the atria. The effective contraction of the atria is lost, allowing blood to pool and clot, leading to stroke if untreated.
Battery capacity: Given by the voltage and the current delivery. The voltage is a result of the battery chemistry, and current delivery (current x time) is measured in ampere hours and is related to battery size.
CMOS circuit: Abbreviation for complementary metallic oxide semiconductor, which is a form of semiconductor often used in pacemaker technology.
Congestive heart failure: The pathophysiologic state in which an abnormality of cardiac function is responsible for the failure of the heart to pump blood at a rate commensurate with the requirements of the body.
Endocardium: The inner lining of the heart.
Epicardium: The outer lining of the heart.
Hermeticity: The term, as used in the pacemaker industry, refers to a very low rate of helium gas leakage from the sealed pacemaker container. This reduces the chance of fluid intruding into the pacemaker generator and causing damage.
Hypertrophic obstructive cardiomyopathy: A disease of the myocardium characterized by thickening (hypertrophy) of the interventricular septum, resulting in the partial obstruction of blood from the left ventricle.
Minute ventilation: Respiratory rate x tidal volume (the amount of air taken in with each breath) = minute ventilation. This parameter is used as a biologic indicator for rate-adaptive pacing.
Mode: The type of pacemaker response to the patient's intrinsic heartbeat. The three commonly used modes are asynchronous, demand, and triggered.
Programmable: The ability to alter the pacemaker settings noninvasively. A variety of selections exist, each with its own designation.
Rate-adaptive: The ability to change the pacemaker stimulation interval caused by sensing a physiologic function other than the intrinsic atrial rhythm.
Sensitivity: A programmable setting that adjusts the reference voltage to which signals entering the sensing circuit are compared for filtering.
Stimulation threshold: The minimum output energy required to consistently "capture" (cause depolarization) of the heart.

References

Andersen HR, Thuesen L, Bagger JP, et al. 1993. Atrial versus ventricular pacing in sick sinus syndrome: A prospective randomized trial in 225 consecutive patients. *Eur Heart J* 14(abstr suppl):252.

Bernstein AD, Camm AJ, Fletcher RD, et al. 1987. The NASPE/BPEG generic pacemaker code for antibradyarrhythmia and adaptive-rate pacing and antitachyarrhythmia devices. *PACE* 10:794.

Committee on Pacemaker Implantation. 1991. Guidelines for implantation of cardiac pacemakers and antiarrhythmic devices. *J Am Coll Cardiol* 18(1):1.

Further Information

A good basic introduction to pacing from a clinical perspective is the third edition of *A Practical Guide to Cardiac Pacing,* by H.Weston Moses, Joel Schneider, Brian Miller, and George Taylor (Little, Brown, 1991).

Cardiac Pacing (Blackwell Scientific, 1992), edited by Kenneth Ellenbogen, is an excellent intermediate treatment of pacing. The treatments of timing cycles and troubleshooting are especially good.

In-depth discussion of a wide range of pacing topics is provided by the third edition of *A Practice of Cardiac Pacing,* by Seymour Furman, David Hayes, and David Holmes (Futura, 1993), and by *New Perspectives in Cardiac Pacing 3,* edited by Serge Barold and Jacques Mugica (Futura, 1993).

Detailed treatment of rate-responsive pacing is given in *Rate-Adaptive Cardiac Pacing: Single and Dual Chamberby* Chu-Pak Lau (Futura, 1993), and in *Rate-Adaptive Pacing,* edited by David Benditt (Blackwell Scientific, 1993).

The Foundations of Cardiac Pacing, Part I, by Richard Sutton and Ivan Bourgeois (Futura, 1991) contains excellent illustrations of implantation techniques.

Readers seeking a historical perspective may wish to consult “Pacemakers, Pastmakers, and the Paced: An Informal History from A to Z,” by Dwight Harken in the July/August 1991 issue of *Biomedical Instrumentation and Technology*.

PACE is the official journal of the North American Society of Pacing and Electrophysiology (NASPE) and of the International Cardiac Pacing and Electrophysiology Society. It is published monthly by Futura Publishing (135 Bedford Road, PO Box 418, Armonk, NY 10504 USA).

22
Implantable Stimulators for Neuromuscular Control

Primož Strojnik
Case Western Reserve University

P.Hunter Peckham
Case Western Reserve University
and Veterans Affairs Medical Center

22.1 Functional Electrical Stimulation

Implantable stimulators for neuromuscular control are the technologically most advanced versions of functional electrical stimulators. Their function is to generate contraction of muscles, which cannot be controlled volitionally because of the damage or dysfunction in the neural paths of the central nervous system (CNS). Their operation is based on the electrical nature of conducting information within nerve fibers, from the neuron cell body (soma), along the axon, where a travelling action potential is the carrier of excitation. While the action potential is naturally generated chemically in the head of the axon, it may also be generated artificially by depolarizing the neuron membrane with an electrical pulse. A train of electrical impulses with certain amplitude, width, and repetition rate, applied to a muscle innervating nerve (a motor neuron) will cause the muscle to contract, very much like in natural excitation. Similarly, a train of electrical pulses applied to the muscular tissue close to the motor point will cause muscle contraction by stimulating the muscle through the neural structures at the motor point.

22.2 Technology for Delivering Stimulation Pulses to Excitable Tissue

A practical system used to stimulate a nerve consists of three components: (1) *a pulse generator* to generate a train of pulses capable of depolarizing the nerve, (2) a *lead wire*, the function of which is to deliver the pulses to the stimulation site, and (3) an *electrode*, which delivers the stimulation pulses to the excitable tissue in a safe and efficient manner.

In terms of location of the above three components of an electrical stimulator, stimulation technology can be described in the following terms:

0-8493-1813-0/03/$0.00+$1.50

Surface or transcutaneous stimulation, where all three components are outside the body and the electrodes are placed on the skin above or near the motor point of the muscle to be stimulated. This method has been used extensively in medical rehabilitation of nerve and muscle. Therapeutically, it has been used to prevent atrophy of paralyzed muscles, to condition paralyzed muscles before the application of functional stimulation, and to generally increase the muscle bulk. As a functional tool, it has been used in rehabilitation of plegic and paretic patients. Surface systems for functional stimulation have been developed to correct drop-foot condition in hemiplegic individuals [Liberson, 1961], for hand control [Rebersek, 1973], and for standing and stepping in individuals with paralysis of the lower extremities [Kralj and Bajd, 1989]. This fundamental technology was commercialized by Sigmedics, Inc. [Graupe, 1998]. The inability of surface stimulation to reliably excite the underlying tissue in a repeatable manner and to selectively stimulate deep muscles has limited the clinical applicability of surface stimulation.

Percutaneous stimulation employs electrodes that are positioned inside the body close to the structures to be stimulated. Their lead wires permanently penetrate the skin to be connected to the external pulse generator. State of the art embodiments of percutaneous electrodes utilize a small-diameter insulated stainless steel lead that is passed through the skin. The electrode structure is formed by removal of the insulation from the lead and subsequent modification to ensure stability within the tissue. This modification includes forming barbs or similar anchoring mechanisms. The percutaneous electrode is implanted using a hypodermic needle as a trochar for introduction. As the needle is withdrawn, the anchor at the electrode tip is engaged into the surrounding tissue and remains in the tissue. A connector at the skin surface, next to the skin penetration point, joins the percutaneous electrode lead to the hardwired external stimulator. The penetration site has to be maintained and care must be taken to avoid physical damage of the lead wires. In the past, this technology has helped develop the existing implantable systems, and it may be used for short and long term, albeit not permanent, stimulation applications [Memberg, 1993; Marsolais, 1986].

The term *implantable stimulation* refers to stimulation systems in which all three components, pulse generator, lead wires, and electrodes, are permanently surgically implanted into the body and the skin is solidly closed after the implantation procedure. Any interaction between the implantable part and the outside world is performed using telemetry principles in a contactless fashion. This chapter is focused on implantable neuromuscular stimulators, which will be discussed in more detail.

22.3 Stimulation Parameters

In functional electrical stimulation, the typical stimulation waveform is a train of rectangular pulses. This shape is used because of its effectiveness as well as relative ease of generation. All three parameters of a stimulation train, i.e., frequency, amplitude, and pulse width, have an effect on muscle contraction. Generally, the stimulation frequency is kept as low as possible, to prevent muscle fatigue and to conserve stimulation energy. The determining factor is the muscle fusion frequency at which a smooth muscle response is obtained. This frequency varies; however, it can be as low as 12 to 14 Hz and

as high as 50 Hz. In most cases, the stimulation frequency is kept constant for a certain application. This is true both for surface as well as implanted electrodes.

With surface electrodes, the common way of modulating muscle force is by varying the stimulation pulse amplitude at a constant frequency and pulse width. The stimulation amplitudes may be as low as 25V at 200 μs for the stimulation of the peroneal nerve and as high as 120V or more at 300 us for activation of large muscles such as the gluteus maximus.

In implantable stimulators and electrodes, the stimulation parameters greatly depend on the implantation site. When the electrodes are positioned on or around the target nerve, the stimulation amplitudes are on the order of a few milliamperes or less. Electrodes positioned on the muscle surface (epimysial electrodes) or in the muscle itself (intramuscular electrodes), employ up to ten times higher amplitudes. For muscle force control, implantable stimulators rely either on pulse-width modulation or amplitude modulation. For example, in upper extremity applications, the current amplitude is usually a fixed paramter set to 16 or 20 mA, while the muscle force is modulated with pulse widths within 0 to 200 us.

22.4 Implantable Neuromuscular Stimulators

Implantable stimulation systems use an encapsulated pulse generator that is surgically implanted and has subcutaneous leads that terminate at electrodes on or near the desired nerves. In low-power consumption applications such as the cardiac pacemaker, a primary battery power source is included in the pulse generator case. When the battery is close to depletion, the pulse generator has to be surgically replaced.

Most implantable systems for neuromuscular application consist of an external and an implanted component. Between the two, an inductive radio-frequency link is established, consisting of two tightly coupled resonant coils. The link allows transmission of power and information, through the skin, from the external device to the implanted pulse generator. In more advanced systems, a back-telemetry link is also established, allowing transmission of data outward, from the implanted to the external component.

Ideally, implantable stimulators for neuromuscular control would be standalone, totally implanted devices with an internal power source and integrated sensors detecting desired movements from the motor cortex and delivering stimulation sequences to appropriate muscles, thus bypassing the neural damage. At the present developmental stage, they still need a control source and an external controller to provide power and stimulation information. The control source may be either operator driven, controlled by the user, or triggered by an event such as the heel-strike phase of the gait cycle. Figure 22.1 depicts a neuromuscular prosthesis developed at the Case Western Reserve University (CWRU) and Cleveland Veterans Affairs Medical Center for the restoration of hand functions using an implantable

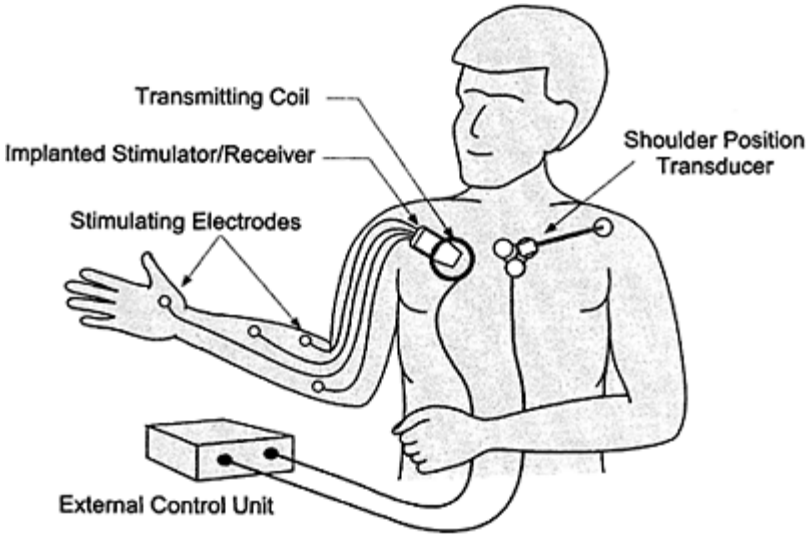


FIGURE 22.1 Implanted FES hand-grasp system.

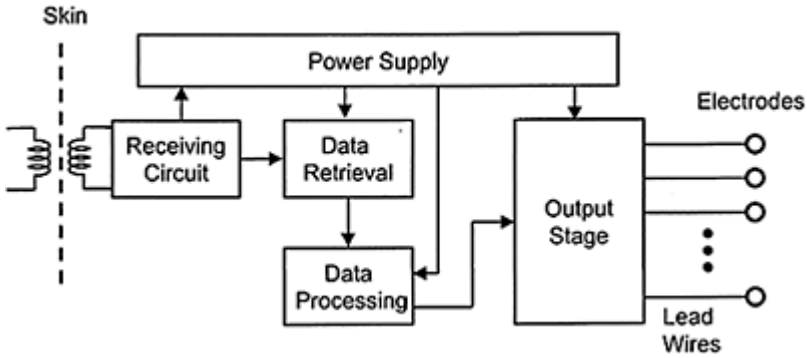


FIGURE 22.2 Block diagram of an implantable neuromuscular stimulator.

neuromuscular stimulator. In this application, the patient uses the shoulder motion to control opening and closing of the hand.

The internal electronic structure of an implantable neuromuscular stimulator is shown in Fig. 22.2. It consists of receiving and data retrieval circuits, power supply, data processing circuits, and output stages.

Receiving Circuit

The stimulator's receiving circuit is an LC circuit tuned to the resonating frequency of the external transmitter, followed by a rectifier. Its task is to provide the raw DC power from the received **rf** signal and at the same time allow extraction of stimulation information embedded in the **rf** carrier. There are various encoding schemes allowing simultaneous transmission of power and information into an implantable electronic device. They include amplitude and frequency modulation with different modulation indexes as well as different versions of digital encoding such as Manchester encoding, where the information is hidden in a logic value transition position rather than the logic value itself. Synchronous and asynchronous clock signals may be extracted from the modulated carrier to drive the implant's logic circuits.

The use of radiofrequency transmission for medical devices is regulated and in most countries limited to certain frequencies and radiation powers. (In the U.S., the use of the **rf** space is regulated by the Federal Communication Commission [FCC]). Limited **rf** transmission powers as well as conservation of power in battery operated external controllers dictate high coupling efficiencies between the transmitting and receiving antennas. Optimal coupling parameters cannot be uniformly defined; they depend on application particularities and design strategies.

Power Supply

The amount of power delivered into an implanted electronic package depends on the coupling between the transmitting and the receiving coil. The coupling is dependent on the distance as well as the alignment between the coils. The power supply circuits must compensate for the variations in distance for different users as well as for the alignment variations due to skin movements and consequent changes in relative coil-to-coil position during daily usage. The power dissipated on power supply circuits must not raise the overall implant case temperature.

In implantable stimulators that require stimulation voltages in excess of the electronics power-supply voltages (20 to 30 V), the stimulation voltage can be provided directly through the receiving coil. In that case, voltage regulators must be used to provide the electronics supply voltage (usually 5 V), which heavily taxes the external power transmitter and increases the implant internal power dissipation.

Data Retrieval

Data retrieval technique depends on the data-encoding scheme and is closely related to power supply circuits and implant power consumption. Most commonly, amplitude modulation is used to encode the in-going data stream. As the high quality factor of resonant LC circuits increases the efficiency of power transmission, it also effectively reduces the transmission bandwidth and therefore the transmission data rate. Also, high-quality circuits are difficult to amplitude modulate since they tend to continue oscillating even with power removed. This has to be taken into account when designing the communication link in particular for the start-up situation when the implanted device does not use the power for stimulation and therefore loads the transmitter side less

heavily, resulting in narrower and higher resonant curves. The load on the receiving coil may also affect the low-pass filtering of the received **rf** signal.

Modulation index (m) or depth of modulation affects the overall energy transfer into the implant. At a given **rf** signal amplitude, less energy is transferred into the implanted device when 100% modulation is used (m=1) as compared to 10% modulation (m=0.053). However, retrieval of 100% modulated signal is much easier than retrieval of a 10% modulated signal.

Data Processing

Once the information signal has been satisfactorily retrieved and reconstructed into logic voltage levels, it is ready for logic processing. For synchronous data processing a clock signal is required. It can be generated locally within the implant device, reconstructed from the incoming data stream, or can be derived from the **rf** carrier. A crystal has to be used with a local oscillator to assure stable clock frequency. Local oscillator allows for asynchronous data transmission. Synchronous transmission is best achieved using Manchester data encoding. Decoding of Manchester-encoded data recovers the original clock signal, which was used during data encoding. Another method is using the downscaled **rf** carrier signal as the clock source. In this case, the information signal has to be synchronized with the **rf** carrier. Of course, 100% modulation scheme cannot be used with carrier-based clock signal. Complex command structure used in multichannel stimulators requires intensive data decoding and processing and consequently extensive electronic circuitry. Custom-made, application specific circuits (ASIC) are commonly used to minimize the space requirements and optimize the circuit performance.

Output Stage

The output stage forms stimulation pulses and defines their electrical characteristics. Even though a mere rectangular pulse can depolarize a nervous membrane, such pulses are not used in clinical practice due to their noxious effect on the tissue and stimulating electrodes. These effects can be significantly reduced by charge-balanced stimulating pulses where the cathodic stimulation pulse is followed by an anodic pulse containing the same electrical charge, which reverses the electrochemical effects of the cathodic pulse. Charge-balanced waveforms can be assured by capacitive coupling between the pulse generator and stimulation electrodes. Charge-balanced stimulation pulses include symmetrical and asymmetrical waveforms with anodic phase immediately following the cathodic pulse or being delayed by a short, 20 to 60-μs interval.

The output stages of most implantable neuromuscular stimulators have constant current characteristics, meaning that the output current is independent on the electrode or tissue impedance. Practically, the constant current characteristics ensure that the same current flows through the excitable tissues regardless of the changes that may occur on the electrode-tissue interface, such as the growth of fibrous tissue around the electrodes. Constant current output stage can deliver constant current only within the supply voltage—compliance voltage. In neuromuscular stimulation, with the electrode impedance being on the order of 1 kΩ, and the stimulating currents in the order of 20 mA, the compliance voltage must be above 20 V. Considering the voltage drops and

losses across electronic components, the compliance voltage of the output stage may have to be as high as 33 V.

The stimulus may be applied through either monopolar or bipolar electrodes. The monopolar electrode is one in which a single active electrode is placed near the excitable nerve and the return electrode is placed remotely, generally at the implantable unit itself. Bipolar electrodes are placed at the stimulation site, thus limiting the current paths to the area between the electrodes. Generally, in monopolar stimulation the active electrode is much smaller than the return electrode, while bipolar electrodes are the same size.

22.5 Packaging of Implantable Electronics

Electronic circuits must be protected from the harsh environment of the human body. The packaging of implantable electronics uses various materials, including polymers, metals, and ceramics. The encapsulation method depends somewhat on the electronic circuit technology. Older devices may still use discrete components in a classical form, such as leaded transistors and resistors. The newer designs, depending on the sophistication of the implanted device, may employ application-specific integrated circuits (ASICs) and thick film hybrid circuitry for their implementation. Such circuits place considerable requirements for hermeticity and protection on the implanted circuit packaging.

Epoxy encapsulation was the original choice of designers of implantable neuromuscular stimulators. It has been successfully used with relatively simple circuits using discrete, low-impedance components. With epoxy encapsulation, the receiving coil is placed around the circuitry to be “potted” in a mold, which gives the implant the final shape. Additionally, the epoxy body is coated with silicone rubber that improves the biocompatibility of the package. Polymers do not provide an impermeable barrier and therefore cannot be used for encapsulation of high-density, high-impedance electronic circuits. The moisture ingress ultimately will reach the electronic components, and surface ions can allow electric shorting and degradation of leakage-sensitive circuitry and subsequent failure.

Hermetic packaging provides the implant electronic circuitry with a long-term protection from the ingress of body fluids. Materials that provide hermetic barriers are metals, ceramics, and glasses. Metallic packaging generally uses a titanium capsule machined from a solid piece of metal or deep-drawn from a piece of sheet metal. Electrical signals, such as power and stimulation, enter and exit the package through hermetic feedthroughs, which are hermetically welded onto the package walls. The feedthrough assembly utilizes a ceramic or glass insulator to allow one or more wires to exit the package without contact with the package itself. During the assembly procedures, the electronic circuitry is placed in the package and connected internally to the feedthroughs, and the package is then welded closed. Tungsten Inert Gas (TIG), electron beam, or laser welding equipment is used for the final closure. Assuming integrity of all components, hermeticity with this package is ensured. This integrity can be checked by detecting gas leakage from the capsule. Metallic packaging requires that the receiving coil be placed outside the package to avoid significant loss of **rf** signal or power, thus requiring additional space within the body to accommodate the volume of the entire implant. Generally, the hermetic package and the receiving antenna are jointly imbedded

in an epoxy encapsulant, which provides electric isolation for the metallic antenna and stabilizes the entire implant assembly. Figure 22.3 shows such an implantable stimulator designed and made by the CWRU/ Veterans Administration Program. The hermetic package is open, displaying the electronic hybrid circuit. More recently, alumina-based ceramic packages have been developed that allow hermetic sealing of the electronic circuitry together with enclosure of the receiving coil [Strojnik, 1994]. This is possible due to the **rf** transparency of ceramics. The impact of this type of enclosure is still not fully investigated. The advantage of this approach is that the volume of the implant can be reduced, thus minimizing the biologic response, which is a function of volume. Yet, an unexplored issue of this packaging method is the effect of powerful electromagnetic fields on the implant circuits, lacking the protection of the metal enclosure. This is a particular concern with high-gain (EMG, ENG, or EKG sensing) amplifiers, which in the future may be included in the implant package as part of back-telemetry circuits. Physical strength of ceramic packages and their resistance to impact will also require future investigation.

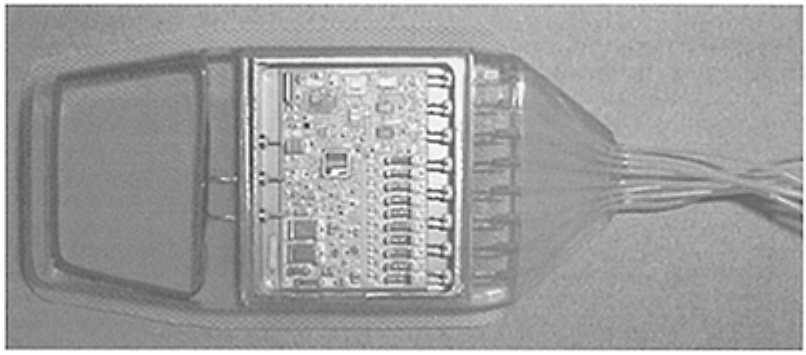


FIGURE 22.3 Photograph of a multichannel implantable stimulator telemeter. Hybrid circuit in titanium package is shown exposed. Receiving coil (left) is imbedded in epoxy resin together with titanium case. Double feedthroughs are seen penetrating titanium capsule wall on the right.

22.6 Leads and Electrodes

Leads connect the pulse generator to the electrodes. They must be sufficiently flexible to move across the joints while at the same time sufficiently sturdy to last for the decades of the intended life of the device. They must also be stretchable to allow change of distance between the pulse generator and the electrodes, associated with body movements. Ability

to flex and to stretch is achieved by coiling the lead conductor into a helix and inserting the helix into a small-diameter silicone tubing. This way, both flexing movements and stretching forces exerted on the lead are attenuated, while translated into torsion movements and forces exerted on the coiled conductor. Using multistrand rather than solid conductors further enhances the longevity. Several individually insulated multi-strand conductors can be coiled together, thus forming a multiple conductor lead wire. Most lead configurations include a connector at some point between the implant and the terminal electrode, allowing for replacement of the implanted receiver or leads in the event of failure. The connectors used have been either single pin in-line connectors located somewhere along the lead length or a multiport/multilead connector at the implant itself. Materials used for lead wires are stainless steels, MP35N (Co, Cr, Ni alloy), and noble metals and their alloys.

Electrodes deliver electrical charge to the stimulated tissues. Those placed on the muscle surface are called epimysial, while those inserted into the muscles are called intramuscular. Nerve stimulating electrodes are called epineural when placed against the nerve, or cuff electrodes when they encircle the nerve. Nerve electrodes may embrace the nerve in a spiral manner individually, or in an array configuration. Some implantable stimulation systems merely use exposed lead-wire conductor sutured to the epineurium as the electrode. Generally, nerve electrodes require approximately one-tenth of the energy for muscle activation as compared to muscle electrodes. However, they require more extensive surgery and may be less selective, but the potential for neural damage is greater than, for example, nerve encircling electrodes.

Electrodes are made of corrosion resistant materials, such as noble metals (platinum or iridium) and their alloys. For example, a platinum—iridium alloy consisting of 10% iridium and 90% platinum is commonly used as an electrode material. Epimysial electrodes developed at CWRU use Ø4 mm Pt90Ir10 discs placed on Dacron reinforced silicone backing. CWRU intramuscular electrodes employ a stainless steel lead-wire with the distal end de-insulated and configured into an electrode tip. A small, umbrella-like anchoring barb is attached to it. With this arrangement, the diameter of the electrode tip does not differ much from the lead wire diameter and this electrode can be introduced into a deep muscle with a trochar-like insertion tool. Figure 22.4 shows enlarged views of these electrodes.

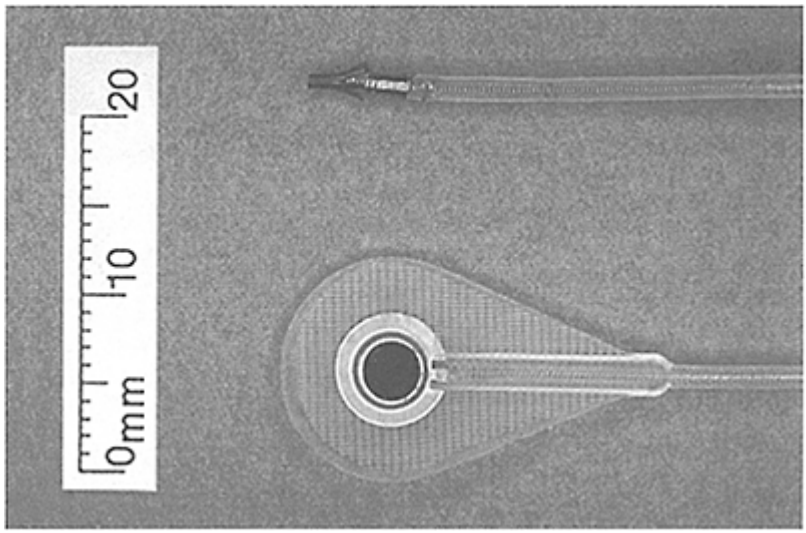


FIGURE 22.4 Implantable electrodes with attached lead wires. Intramuscular electrode (top) has stainless steel tip and anchoring barbs. Epimysial electrode has PtIr disk in the center and is backed by silicone-impregnated Dacron mesh.

22.7 Safety Issues of Implantable Stimulators

The targeted lifetime of implantable stimulators for neuromuscular control is the lifetime of their users, which is measured in tens of years. Resistance to premature failure must be assured by manu-facturing processes and testing procedures. Appropriate materials must be selected that will withstand the working environment. Protection against mechanical and electrical hazards that may be encoun-tered during the device lifetime must be incorporated in the design. Various procedures are followed and rigorous tests must be performed during and after manufacturing to assure the quality and reliability of the device.

- **Manufacturing and testing**—Production of implantable electronic circuits and their encapsu-lation in many instances falls under the standards governing production and encapsulation of integrated circuits. To minimize the possibility of failure, the implantable electronic devices are manufactured in controlled clean-room environments, using high-quality components and strictly defined manufacturing procedures. Finished devices are submitted to rigorous testing before being released for implantation. Also, many tests are carried out during the manufacturing process

itself. To assure maximum reliability and product confidence, methods, tests, and procedures defined by military standards, such as MIL-STD-883, are followed.

- **Biocompatibility**—Since the implantable stimulators operate surgically implanted in living tissue, an important part of their design has to be dedicated to biocompatibility, i.e., their ability to dwell in living tissue without disrupting the tissue in its functions, creating adverse tissue response, or changing its own properties due to the tissue environment. Elements of biocompat-ibility include tissue reaction to materials, shape, and size, as well as electrochemical reactions on stimulation electrodes. There are known biomaterials used in the making of implantable stimu-lators. They include stainless steels, titanium and tantalum, noble metals such as platinum and iridium, as well as implantable grades of selected epoxy and silicone-based materials.
- **Susceptibility to electromagnetic interference (EMI) and electrostatic discharge (ESD)**—Electromagnetic fields can disrupt the operation of electronic devices, which may be lethal in situations with life-support systems, but they may also impose risk and danger to users of neuromuscular stimulators. Emissions of EMI may come from outside sources; however, the external control unit is also a source of electromagnetic radiation. Electrostatic discharge shocks are not uncommon during the dry winter season. These shocks may reach voltages as high as 15 kV and more. Sensitive electronic components can easily be damaged by these shocks unless protective design measures are taken. The electronic circuitry in implantable stimulators is generally protected by the metal case. However, the circuitry can be damaged through the feedthroughs either by handling or during the implantation procedure by the electrocautery equipment. ESD damage may happen even after implantation when long lead-wires are utilized. There are no standards directed specifically toward implantable electronic devices. The general standards put in place for electromedical equipment by the International Electrotechnical Commission provide guidance. The specifications require survival after 3 kV and 8 kV ESD discharges on all conductive and nonconductive accessible parts, respectively.

22.8 Implantable Stimulators in Clinical Use

Peripheral Nerve Stimulators

- **Manipulation**—Control of complex functions for movement, such as hand control, requires the use of many channels of stimulation. At the Case Western Reserve University and Cleveland VAMC, an eight-channel stimulator has been developed for grasp and release [Smith, 1987]. This system uses eight channels of stimulation and a titanium-packaged, thick-film hybrid circuit as the pulse generator. The implant is distributed by the Neurocontrol Corporation (Cleveland, OH) under the name of Freehand®. It has been implanted in approximately 150 patients in the U.S., Europe, Asia, and Australia. The implant is controlled by a dual-microprocessor external unit carried by the patient with an input control signal provided by the user's remaining volitional movement. Activation of the muscles provides two primary grasp patterns and allows the person to achieve functional performance that exceeds his or her

capabilities without the use of the implanted system. This system received premarket approval from the FDA in 1998.

- **Locomotion**—The first implantable stimulators were designed and implanted for the correction of the foot-drop condition in hemiplegic patients. Medtronic's Neuromuscular Assist (NMA) device consisted of an **rf** receiver implanted in the inner thigh and connected to a cuff electrode embracing the peroneal nerve just beneath the head of fibula at the knee [McNeal, 1977; Waters 1984]. The Ljubljana peroneal implant had two versions [Vavken, 1976; Strojnik, 1987] with the common feature that the implant-rf receiver was small enough to be implanted next to the peroneal nerve in the fossa poplitea region. Epineural stimulating electrodes were an integral part of the implant. This feature and the comparatively small size make the Ljubljana implant a precursor of the microstimulators described in Section 22.9. Both NMA and the Ljubljana implants were triggered and synchronized with gait by a heel switch.

The same implant used for hand control and developed by the CWRU has also been implanted in the lower extremity musculature to assist incomplete quadriplegics in standing and transfer operations [Triolo, 1996]. Since the design of the implant is completely transparent, it can generate any stimulation sequence requested by the external controller. For locomotion and transfer-related tasks, stimulation sequences are preprogrammed for individual users and activated by the user by means of pushbuttons. The implant (two in some applications) is surgically positioned in the lower abdominal region. Locomotion application uses the same electrodes as the manipulation system; however, the lead wires have to be somewhat longer.

- **Respiration**—Respiratory control systems involve a two-channel implantable stimulator with electrodes applied bilaterally to the phrenic nerve. Most of the devices in clinical use were developed by Avery Laboratories (Dobelle Institute) and employed discrete circuitry with epoxy encapsula-tion of the implant and a nerve cuff electrode. Approximately 1000 of these devices have been implanted in patients with respiratory disorders such as high-level tetraplegia [Glenn, 1986]. Activation of the phrenic nerve results in contraction of each hemidiaphragm in response to electrical stimulation. In order to minimize damage to the diaphragms during chronic use, alternation of the diaphragms has been employed, in which one hemidiaphragm will be activated for several hours followed by the second. A review of existing systems was given by Creasy et al. [1996]. Astrotech of Finland also recently introduced a phrenic stimulator. More recently, DiMarco [1997] has investigated use of CNS activation of a respiratory center to provide augmented breathing.
- **Urinary control**—Urinary control systems have been developed for persons with spinal cord injury. The most successful of these devices has been developed by Brindley [1982] and is manufactured by Finetech, Ltd. (England). The implanted receiver consists of three separate stimulator devices, each with its own coil and circuitry, encapsulated within a single package. The sacral roots (S2, S3, and S4) are placed within a type of encircling electrode, and stimulation of the proper roots will generate contraction of both the bladder and the external sphincter. Cessation of stimulation results in faster relaxation of the external sphincter than of the bladder wall, which then results in voiding. Repeated trains of pulses applied in this manner will eliminate most urine, with only small residual amounts remaining. Approximately 1500 of these

devices have been implanted around the world. This technology also has received FDA premarket approval and is currently distributed by NeuroControl Corporation.

- **Scoliosis treatment**—Progressive lateral curvature of the adolescent vertebral column with simultaneous rotation is known as idiopathic scoliosis. Electrical stimulation applied to the convex side of the curvature has been used to stop or reduce its progression. Initially **rf** powered stimulators have been replaced by battery-powered, totally implanted devices [Bobechko, 1979; Herbert, 1989]. Stimulation is applied intermittently, stimulation amplitudes are under 10.5V (510Ω), and frequency and pulsewidth are within the usual FES parameter values.

Stimulators of Central Nervous System

Some stimulation systems have electrodes implanted on the surface of the central nervous system or in its deep areas. They do not produce functional movements; however, they "modulate" a pathological motor brain behavior and by that stop unwanted motor activity or abnormality. Therefore, they can be regarded as stimulators for neuromuscular control.

- **Cerebellar stimulation**—Among the earliest stimulators from this category are cerebellar stimulators for control of reduction of effects of cerebral palsy in children. Electrodes are placed on the cerebellar surface with the leads penetrating cranium and dura. The pulse generator is located subcutaneously in the chest area and produces intermittent stimulation bursts. There are about 600 patients using these devices [Davis, 1997].
- **Vagal stimulation**—Intermittent stimulation of the vagus nerve with 30 sec on and five min off has been shown to reduce frequency of epileptic seizures. A pacemaker-like device, developed by Cyberonics, is implanted in the chest area with a bipolar helical electrode wrapped around the left vagus nerve in the neck. The stimulation sequence is programmed (most often parameter settings are 30 Hz, 500 μs, 1.75 mA); however, patients have some control over the device using a hand-held magnet [Terry, 1991]. More than 3000 patients have been implanted with this device, which received the premarketing approval (PMA) from the FDA in 1997.
- **Deep brain stimulation**—Recently, in 1998, an implantable stimulation device (Activa by Medtronic) was approved by the FDA that can dramatically reduce uncontrollable tremor in patients with Parkinson's disease or essential tremor [Koller, 1997]. With this device, an electrode array is placed stereotactically into the ventral intermediate nucleus of thalamic region of the brain. Lead wires again connect the electrodes to a programmable pulse generator implanted in the chest area. Application of high-frequency stimulation (130 Hz, 60 to 210 us, 0.25 to 2.75 V) can immediately suppress the patient's tremor.

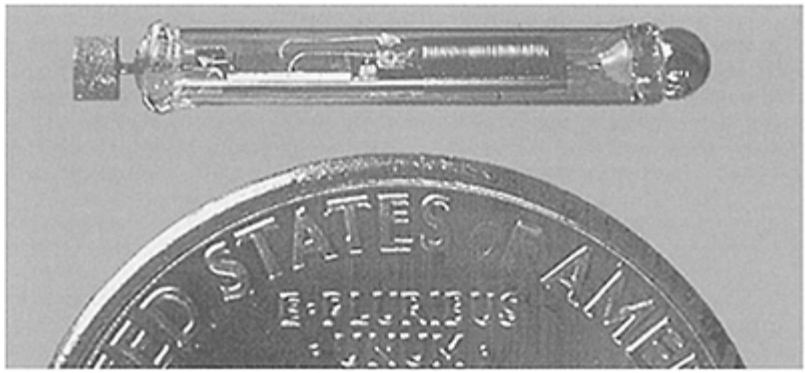


FIGURE 22.5 Microstimulator developed at A.E.Mann Foundation. Dimensions are roughly 2×16 mm. Electrodes at the ends are made of tantalum and iridium, respectively.

22.9 Future of Implantable Electrical Stimulators

Distributed Stimulators

One of the major concerns with multichannel implantable neuromuscular stimulators is the multitude of leads that exit the pulse generator and their management during surgical implantation. Routing of multiple leads virtually increases the implant size and by that the burden that an implant imposes on the tissue. A solution to that may be distributed stimulation systems with a single outside controller and multiple single-channel implantable devices implanted throughout the structures to be stimulated. This concept has been pursued both by the Alfred E. Mann Foundation [Strojnik, 1992; Cameron, 1997] and the University of Michigan [Ziaie, 1997]. Microinjectable stimulator modules have been developed that can be injected into the tissue, into a muscle, or close to a nerve through a lumen of a hypodermic needle. A single external coil can address and activate a number of these devices located within its field, on a pulse-to-pulse basis. A glass-encapsulated microstimulator developed at the AEMF is shown in Fig. 22.5.

Sensing of Implantable Transducer-Generated and Physiological Signals

External command sources such as the shoulder-controlled joystick utilized by the Freehand® system impose additional constraints on the implantable stimulator users, since they have to be donned by an attendant. Permanently implanted control sources make neuro-prosthetic devices much more attractive and easier to use. An implantable joint angle transducer (IJAT) has been developed at the CWRU that consists of a magnet

and an array of magnetic sensors implanted in the distal and the proximal end of a joint, respectively [Smith, 1998]. The sensor is connected to the implantable stimulator package, which provides the power and also transmits the sensor data to the external controller, using a back-telemetry link. Figure 22.6 shows a radiograph of the IJAT implanted in a patient's wrist. Myoelectric signals (MES) from muscles not affected by paralysis are another attractive control source for implantable neuromuscular stimulators. Amplified and bin-integrated EMG signal from uninvolved muscles, such as the sterno-cleido-mastoid muscle, has been shown to contain enough information to control an upper extremity neuroprosthesis [Scott, 1996]. EMG signal is being utilized by a multichannel stimulator-telemeter developed at the CWRU, containing 12 stimulator channels and 2 MES channels integrated into the same platform [Strojnik, 1998].

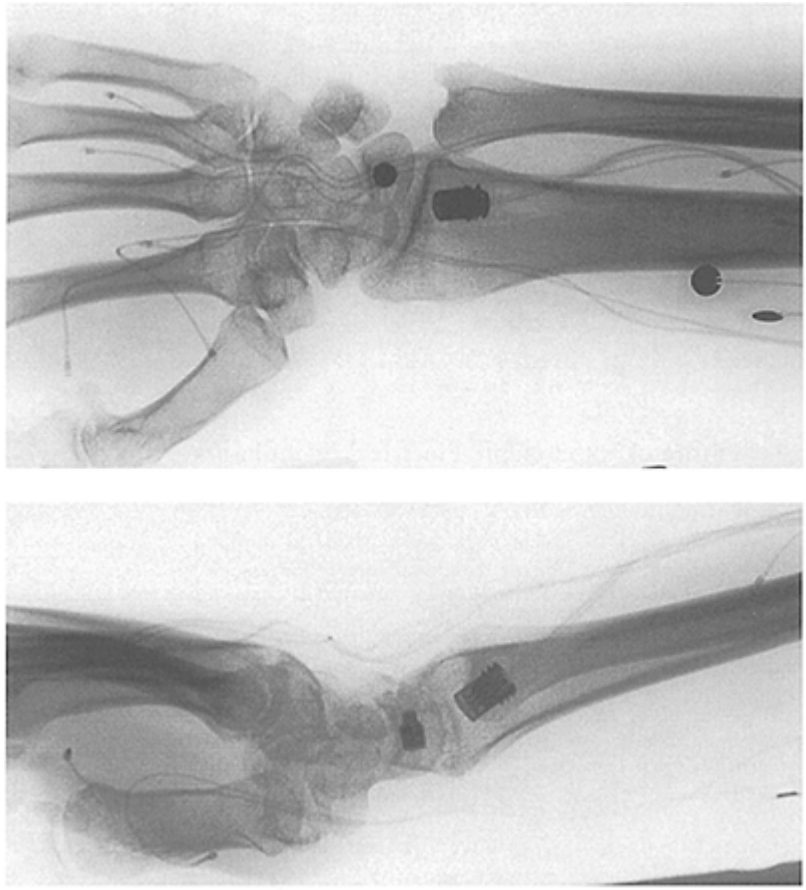


FIGURE 22.6 Radiograph of the joint angle transducer (IJAT) implanted in the wrist. The magnet is implanted in the lunate bone (left) while the

magnetic sensor array is implanted in the radius. Leads going to the implant case can be seen as well as intramuscular and epimysial electrodes with their individual lead wires.

22.10 Summary

Implantable stimulators for neuromuscular control are an important tool in rehabilitation of paralyzed individuals with preserved neuromuscular apparatus, as well as in the treatment of some neurological disorders that result in involuntary motor activity. Their impact on rehabilitation is still in its infancy; however, it is expected to increase with further progress in microelectronics technology, development of smaller and better sensors, and with improvements of advanced materials. Advancements in neuro-physiological science are also expected to bring forward wider utilization of possibilities offered by implantable neuromuscular stimulators.

Defining Terms

Biocompatibility: Ability of a foreign object to coexist in a living tissue.
Electrical stimulation: Diagnostic, therapeutic, and rehabilitational method used to excite motor nerves with the aim of contracting the appropriate muscles and obtain limb movement.
EMG activity: Muscular electrical activity associated with muscle contraction and production of force.
Feedthrough: Device that allows passage of a conductor through a hermetic barrier.
Hybrid circuit: Electronic circuit combining miniature active and passive components on a single ceramic substrate.
Implantable stimulator: Biocompatible electronic stimulator designed for surgical implantation and operation in a living tissue.
Lead wire: Flexible and strong insulated conductor connecting pulse generator to stimulating electrodes.
Paralysis: Loss of power of voluntary movement in a muscle through injury to or disease to its nerve supply.
Stimulating electrode: Conductive device that transfers stimulating current to a living tissue. On its surface, the electric charge carriers change from electrons to ions or vice versa.
rf-radiofrequency: Pertaining to electromagnetic propagation of power and signal in frequencies above those used in electrical power distribution.

References

Bobechko WP, Herbert MA, Friedman HG. 1979. Electrospinal instrumentation for scoliosis: Current status. *Orthop Clin North Am* 10(4):927.

Brindley GS, Polkey CE, Rushton DN. 1982. Sacral anterior root stimulators for bladder control in paraplegia. *Paraplegia* 20(6):365.

Cameron T, Loeb GE, Peck RA, Schulman JH, Strojnik P, Troyk PR. 1997. Micromodular implants to provide electrical stimulation of paralyzed muscles and limbs. *IEEE Trans Biomed Eng* 44(9):781.

Creasey G, Elefteriades J, DiMarco A, Talonen P, Bijak M, Girsch W, Kantor C. 1996. Electrical stimulation to restore respiration. *J Rehab Res Dev* 33(2):123.

Davis R.: Cerebellar stimulation for movement disorders. 1997. In PL Gildenberg and RR Tasker (eds), *Textbook of Stereotactic and Functional Neurosurgery*, McGraw-Hill, New York.

DiMarco AF, Romaniuk JR, Kowalski KE, Supinski GS. 1997. Efficacy of combined inspiratory intercostal and expiratory muscle pacing to maintain artificial ventilation. *Am J Respir Crit Care Med* 156(1):122.

Glenn WW, Phelps ML, Elefteriades JA, Dentz B, Hogan JF. 1986. Twenty years of experience in phrenic nerve stimulation to pace the diaphragm pacing. *Clin Electrophysiol* 9(6 Pt 1):780.

Graupe D, Kohn KH. 1998. Functional neuromuscular stimulator for short-distance ambulation by certain thoracic-level spinal-cord-injured paraplegics. *Surg Neurol* 50(3):202.

Herbert MA, Bobechko WP. 1989. Scoliosis treatment in children using a programmable, totally implantable muscle stimulator (ESI). *IEEE Trans Biomed Eng* 36(7):801.

Koller W, Pahwa R, Busenbark K, Hubble J, Wilkinson S, Lang A, Tuite P, Sime E, Lazano A, Hauser R, Malapira T, Smith D, Tarsy D, Miyawaki E, Norregaard T, Kormos T, Olanow CW. 1997. High-frequency unilateral thalamic stimulation in the treatment of essential and parkinsonian tremor. *Ann Neurol* 42(3):292.

Kralj A, Bajd T. 1989. *Functional Electrical Stimulation: Standing and Walking after Spinal Cord Injury*, CRC Press, Inc., Boca Raton, FL.

Liberson WT, Holmquest, HJ Scot D, Dow M. 1961. Functional electrotherapy: Stimulation of the peroneal nerve synchronized with the swing phase of the gait of hemiplegic patients. *Arch Phys Med & Reb* 42:101.

Marsolais EB, Kobetic R. 1986. Implantation techniques and experience with percutaneous intramuscular electrodes in lower extremities. *J Rehab Res Dev* 23(3):1.

McNeal DR, Waters R, Reswick J. 1977. Experience with implanted electrodes, *Neurosurgery* 1(2):228.

Memberg, W, Peckham PH, Thorpe GB, Keith, MW, Kicher TP. 1993. An analysis of the reliability of percutaneous intramuscular electrodes in upper extremity FNS applications. *IEEE Trans Biomed Eng*, 1(2):126.

Rebersek S, Vodovnik L. 1973. Proportionally controlled functional electrical stimulation of hand. *Arch Phys Med Rehab* 54:378.

Scott TRD., Peckham PH., Kilgore KL. 1996. Tri-State myoelectric control of bilateral upper extremity neuroprosthesies for tetraplegic individuals. *IEEE Trans Rehab Eng* 2:251.

Smith B, Peckham PH, Keith, MW, Roscoe DD. 1987. An externally powered, multichannel, implantable stimulator for versatile control of paralyzed muscle. *IEEE Trans Biomed Engr* 34:499.

Smith B, Tang, Johnson MW, Pourmehdi S, Gazdik MM, Buckett JR, Peckham PH. 1998. An externally powered, multichannel, implantable stimulator-telemeter for control of paralyzed muscle. *IEEE Trans Biomed Eng* 45(4):463.

Strojnik P, Pourmehdi S, Peckham P. 1998. Incorporating FES control sources into implanatable stimulators. *Proc 6th Vienna International Workshop on Functional Electrostimulation*, Vienna, Austria.

Strojnik P, Meadows P, Schulman JH, Whitmoyer D. 1994. Modification of a cochlear stimulation system for FES applications. *Basic & Applied Myology*, BAM 4(2): 129.
Strojnik P, Acimovic R, Vavken E, Simic V, Stanic U. 1987. Treatment of drop foot using an implantable peroneal underknee stimulator. *Scand J Rehab Med* 19:37.
Strojnik P, Schulman J, Loeb G, Troyk P. 1992. Multichannel FES system with distributed microstimulators. *Proc 14th Ann Int Conf IEEE, MBS*, Paris, pp 1352.
Terry RS, Tarver WB, Zabara J. 1991. The implantable neurocybernetic prosthesis system. *Pacing Clin Electrophysiol* 14(1):86.
Triolo RJ, Bieri C, Uhlir J, Kobetic R, Scheiner A, Marsolais EB. 1996. Implanted functional neuromuscular stimulation systems for individuals with cervical spinal cord injuries: Clinical case reports. *Arch Phys Med Rehabil* 77(11): 1119.
Vavken E, Jeglic A. 1976. Application of an implantable stimulator in the rehabilitation of paraplegic patients. *Int Surg* 61(6–7):335–9.
Waters RL, McNeal DR, Clifford B. 1984. Correction of footdrop in stroke patients via surgically implanted peroneal nerve stimulator. *Acta Orthop Belg* 50(2):285.
Ziaie B, Nardin MD, Coghlan AR, Najafi K. 1997. A single-channel implantable microstimulator for functional neuromuscular stimulation. *IEEE Trans Biomed Eng* 44(10):909.

Further Information

Additional references on early work in FES which augment peer review publications can be found in Proceedings from Conferences in Dubrovnik and Vienna. These are the *External Control of Human Extremities* and the *Vienna International Workshop on Electrostimulation*, respectively.

23
External Defibrillators

Willis A.Tacker
Purdue University

Defibrillators are devices used to supply a strong electric shock (often referred to as a *countershock*) to a patient in an effort to convert excessively fast and ineffective heart rhythm disorders to slower rhythms that allow the heart to pump more blood. External defibrillators have been in common use for many decades for emergency treatment of life-threatening cardiac rhythms as well as for elective treatment of less threatening rapid rhythms. Figure 23.1 shows an external defibrillator.

Cardiac arrest occurs in more than 500,000 people annually in the United States, and more than 70% of the out-of-hospital occurrences are due to cardiac arrhythmia treatable with defibrillators. The most serious arrhythmia treated by a defibrillator is ventricular fibrillation. Without rapid treatment using a defibrillator, ventricular fibrillation causes complete loss of cardiac function and death within minutes. Atrial fibrillation and the more organized rhythms of atrial flutter and ventricular tachycardia can be treated on a less emergent basis. Although they do not cause immediate death, their shortening of the interval between contractions can impair filling of the heart chambers and thus decrease cardiac output. Conventionally, treatment of ventricular fibrillation is called *defibrillation*, whereas treatment of the other tachycardias is called *cardioversion*.

23.1 Mechanism of Fibrillation

Fibrillation is chaotic electric excitation of the myocardium and results in loss of coordinated mechanical contraction characteristic of normal heartbeats. Description of mechanisms leading to, and maintaining, fibrillation and other rhythm disorders are reviewed elsewhere [1] and are beyond the scope of this chapter. In summary, however, these rhythm disorders are commonly held to be a result of reentrant excitation pathways within the heart. The underlying abnormality that leads to the mechanism is the combination of conduction block of cardiac excitation plus rapidly recurring depolarization of the membranes of the cardiac cells. This leads to rapid repetitive propagation of a single excitation wave or of multiple excitatory waves throughout the heart. If the waves are multiple, the rhythm may degrade into total loss of synchronization of cardiac fiber contraction. Without synchronized contraction, the chamber affected will not contract, and this is fatal in the case of ventricular fibrillation. The most common cause of these conditions, and therefore of these rhythm disorders, is cardiac ischemia or infarction as a complication of atherosclerosis. Additional relatively

0-8493-1813-0/03/$0.00+$1.50

common causes include other cardiac disorders, drug toxicity, electrolyte imbalances in the blood, hypothermia, and electric shocks (especially from alternating current).

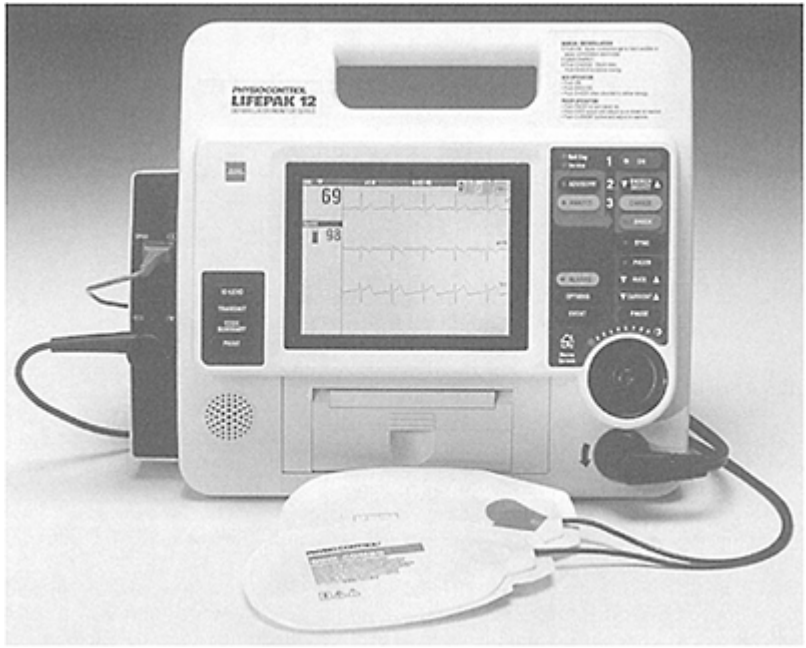


FIGURE 23.1 Photograph of a transchest defibrillator (provided by Physio-Control Corporation with permission).

23.2 Mechanism of Defibrillation

The corrective measure is to extinguish the rapidly occurring waves of excitation by simultaneously depolarizing most of the cardiac cells with a strong electric shock. The cells then can simultaneously repolarize themselves, and thus they will be back in phase with each other.

Despite years of intensive research, there is still no single theory for the mechanism of defibrillation that explains all the phenomena observed. However, it is generally held that the defibrillating shock must be adequately strong and have adequate duration to affect most of the heart cells. In general, longer duration shocks require less current than shorter duration shocks. This relationship is called the strength-duration relationship and is demonstrated by the curve shown in Fig. 23.2. Shocks of strength and duration above and to the right of the current curve (or above the energy curve) have adequate strength to defibrillate, whereas shocks below and to the left do not. From the exponentially decaying current curve an energy curve can also be determined (also shown in Fig. 23.2),

which is high at very short durations due to high current requirements at short durations, but which is also high at longer durations due to additional energy being delivered as the pulse duration is lengthened at nearly constant current. Thus, for most electrical waveforms there is a minimum energy for defibrillation at approximate pulse durations of 3 to 8 ms. A strength-duration charge curve can also be determined as shown in Fig. 23.2, which demonstrates that the minimum charge for defibrillation occurs at the shortest pulse duration tested. Very-short-duration pulses are not used, however, since the high current and voltage required is damaging to the myocardium. It is also important to note that excessively strong or long shocks may cause immediate refibrillation, thus failing to restore the heart function.

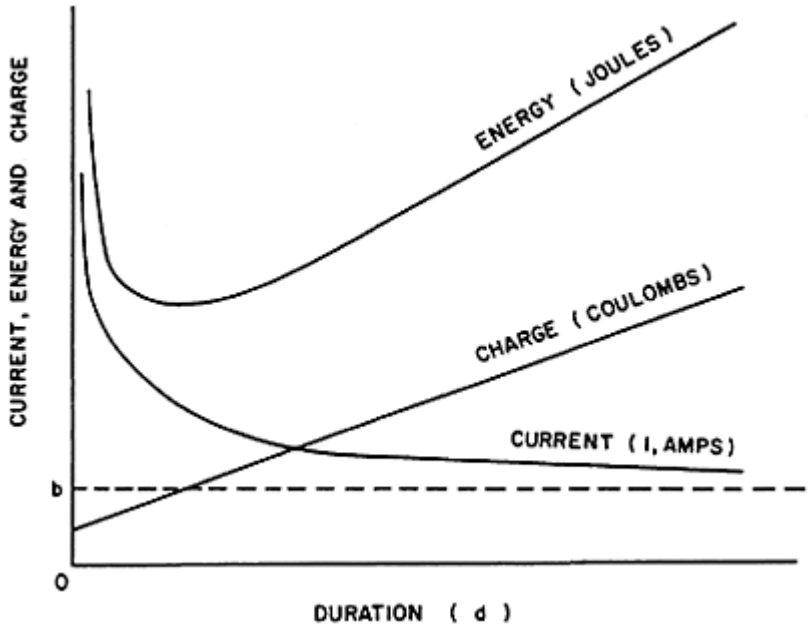


FIGURE 23.2 Strength-duration curves for current, energy, and charge. Adequate current shocks are above and to the right of the current curve. (Modified from Tacker WA, Geddes LA, 1980. *Electrical Defibrillation*, Boca Raton, FL, CRC Press, with permission.)

In practice, for a shock applied to electrodes on the skin surface of the patient's chest, durations are on the order of 3 to 10 milliseconds and have an intensity of a few thousand volts and tens of amperes. The energy delivered to the subject by these shocks is selectable by the operator and is on the order of 50 to 360 joules for most defibrillators.

The exact shock intensity required at a given duration of electric pulse depends on several variables, including the intrinsic characteristics of the patient (such as the underlying disease problem or presence of certain drugs and the length of time the arrhythmia has been present), the techniques for electrode application, and the particular rhythm disorder being treated (more organized rhythms require less energy than disorganized rhythms).

23.3 Clinical Defibrillators

Defibrillator design has resulted from medical and physiologic research and advances in hardware technology. Since is estimated that for each minute that elapses between onset of ventricular fibrillation and the first shock application, an individual's survival rate decreases 10%, it is crucial to respond rapidly. The importance of rapid response led to development of portable, battery-operated defibrillators and more recently to automatic external defibrillators (AEDs) that enable emergency responders to defibrillate with minimal training.

All clinical defibrillators used today store energy in capacitors. Desirable capacitor specifications include small size, light weight, and capability to sustain several thousands of volts and many charge-discharge cycles. Energy storage capacitors account for at least one pound and usually several pounds of defibrillator weight. Energy stored by the capacitor is calculated from

$$W_s = \frac{1}{2} CE^2, \tag{23.1}$$

where W_s=stored energy in joules, C=capacitance in farads, and E=voltage applied to the capacitor. Delivered energy is expressed as

$$W_d = W_S \times \left(\frac{R}{R_i + R} \right), \tag{23.2}$$

where W_d=delivered energy, W_s=stored energy, R=subject resistance, and R_i=device resistance.

Figure 23.3 shows a block diagram for defibrillators. Most have a built-in monitor and synchronizer (dashed lines in Fig. 23.3). Built-in monitoring speeds up diagnosis of potentially fatal arrhythmias, especially when the EGG is monitored through the same electrodes that are used to apply the defibrillating shock. The great preponderance of defibrillators for transchest defibrillation deliver shocks with either a damped sinusoidal waveform produced by discharge of an RCL circuit or a truncated exponential decay waveform (sometimes called trapezoidal). Basic components of exemplary circuits for damped sine waveform and trapezoidal waveform defibrillators are shown in Figs. 23.4 and 23.5. The shape of the

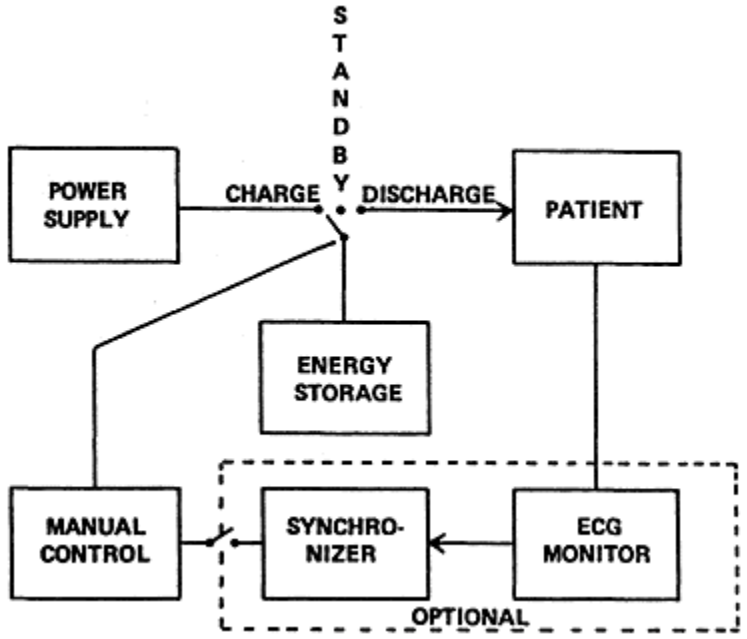


FIGURE 23.3 Block diagram of a typical defibrillator. (From Feinberg B. 1980. *Handbook Series in Clinical Laboratory Science*, Vol 2, Boca Raton, FL, CRC Press, with permission.)

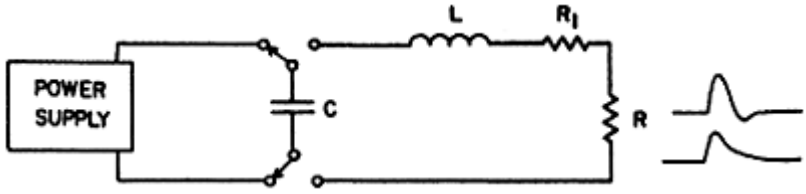


FIGURE 23.4 Resister-capacitor-inductor defibrillator. The patient is represented by *R*. (Modified from Feinberg B. 1980. *Handbook Series in Clinical Laboratory Science*, Vol 2, Boca Raton, FL, CRC Press, with permission.)

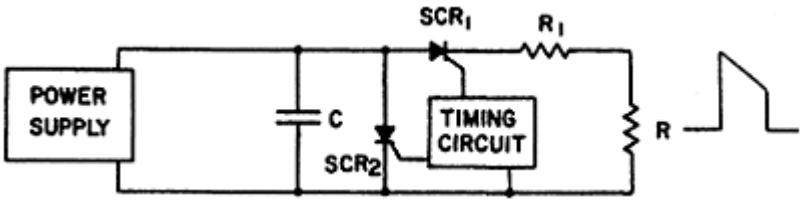


FIGURE 23.5 Trapezoidal wave defibrillator. The patient is represented by *R*. (Modified from Feinberg B. 1980. *Handbook Series in Clinical Laboratory Science,* Vol 2, Boca Raton, FL, CRC Press, with permission.)

waveforms generated by RCL defibrillators depend on the resistance of the patient as well as the energy storage capacitance and resistance and inductance of the inductor. When discharged into a 50-Ω load (to stimulate the patient's resistance), these defibrillators produce either a critically damped sine waveform or a slightly underdamped sine waveform (i.e., having a slight reversal of waveform polarity following the main waveform) into the 50-Ω load.

The exact waveform can be determined by application of Kirkchoff's voltage law to the circuit,

$$L\frac{di}{dt}+\left(R_i+R\right)i+\frac{1}{C}\int idt=0, \tag{23.3}$$

where L=inductance in H, i=instantaneous current in amperes, t=time in seconds, R_i=device resistance, R=subject resistance, and C=capacitance. From this, the second-order differential equation describes the RCL defibrillator:

$$L\frac{d^2i}{dt^2}+\left(R_i+R\right)\frac{di}{dt}+\frac{1}{C}i=0. \tag{23.4}$$

Trapezoidal waveform (actually, these are truncated exponential decay waveform) defibrillators are also used clinically. The circuit diagram in Fig. 23.4 is exemplary of one design for producing such a waveform. Delivered energy calculation for this waveform is expressed as

$$W_d=0.5\,I_i^2\,R\left[\frac{d}{\log_e\left(\frac{I_i}{I_f}\right)}\right]\left[1-\left(\frac{I_f}{I_i}\right)^2\right], \tag{23.5}$$

where W_d=delivered energy, I_i=initial current in amperes, I_f=final current, R=resistance of the patient, and d=pulse duration in seconds. Both RCL and trapezoidal waveforms defibrillate effectively. Implantable defibrillators now use alternative waveforms such as a biphasic exponential decay waveform, in which the polarity of the electrodes is reversed part way through the shock. Use of the biphasic waveform has reduced the shock intensity required for implantable defibrillators but has not yet been extended to transchest use except on an experimental basis.

RCL defibrillators are the most widely available. They store up to about 440 joules and deliver up to about 360 joules into a patient with 50-ohm impedance. Several selectable energy intensities are available, typically from 5 to 360 J, so that pediatric patients, very small patients, or patients with easily converted arrhythmias can be treated with low-intensity shocks. The pulse duration ranges from 3 to 6 ms. Because the resistance *(R)* varies between patients (25 to 150 ohms) and is part of the RCL discharge circuit, the duration and damping of the pulse also varies; increasing patient impedance lengthens and dampens the pulse. Figure 23.6 shows waveforms from RCL defibrillators with critically damped and with underdamped pulses.

23.4 Electrodes

Electrodes for external defibrillation are metal and from 70 to 100 cm^2 in surface area. They must be coupled to the skin with an electrically conductive material to achieve low impedance across the electrode-patient interface. There are two types of electrodes: hand-held (to which a conductive liquid or solid gel is applied) and adhesive, for which an adhesive conducting material holds the electrode in place. Hand-held electrodes are reusable and are pressed against the patient’s chest by the operator during shock delivery. Adhesive electrodes are disposable and are applied to the chest before the shock delivery and left in place for reuse if subsequent shocks are needed. Electrodes are usually applied with both electrodes on the anterior chest as shown in Fig. 23.7 or in anterior-to-posterior (front-to-back) position, as shown in Fig. 23.8.

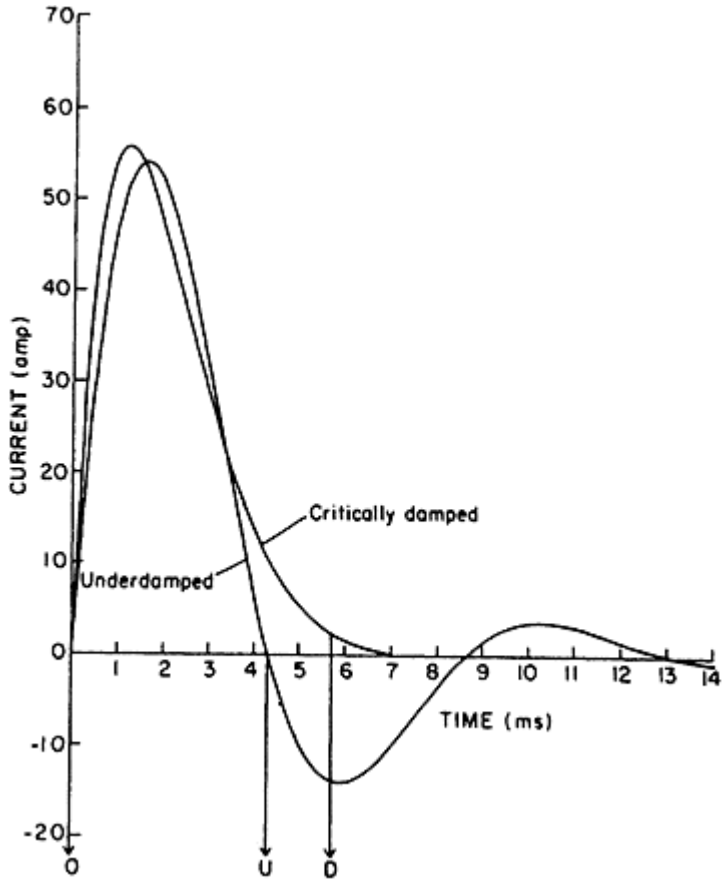


FIGURE 23.6 The damped sine wave. The interval *O-D* represents a duration for the critically and overdamped since waves. By time *D*, more than 99% of the energy has been delivered. *O-U* is taken as the duration for an underdamped sine wave. (Modified from Tacker WA, Geddes LA. 1980. *Electrical Defibrillation*, Boca Raton, FL, CRC Press, with permission.)

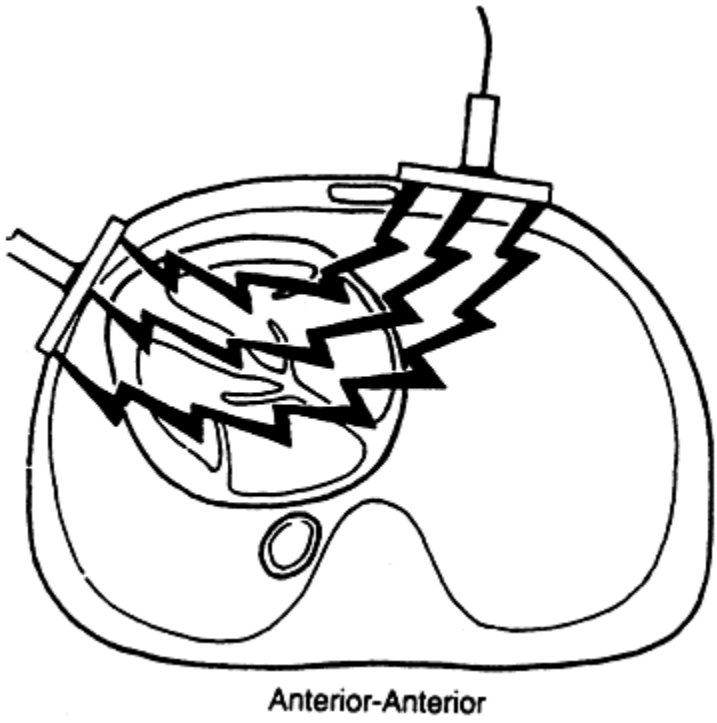


FIGURE 23.7 Cross-sectional view of the chest showing position for standard anterior wall (precordial) electrode placement. Lines of presumed current flow are shown between the electrodes on the skin surface. (Modified from Tacker WA (ed). 1994. *Defibrillation of the Heart: ICDs, AEDs and Manual*, St. Louis, Mosby-Year Book, with permission.)

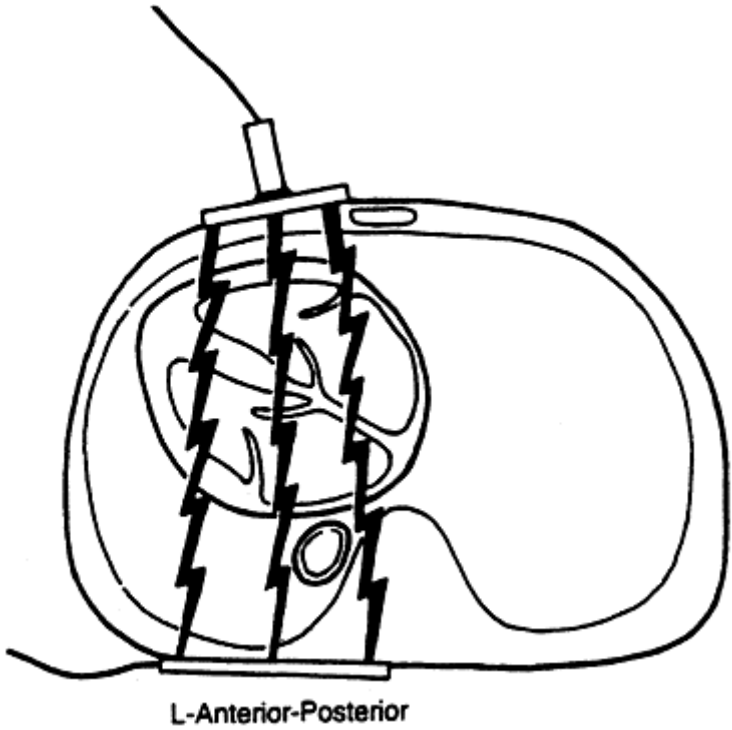


FIGURE 23.8 Cross-sectional view of the chest showing position for front-to-back electrode placement. Lines of presumed current flow are shown between the electrodes on the skin surface. (Modified from Tacker WA (ed). 1994. *Defibrillation of the Heart: ICDs, AEDs and Manual,* St. Louis, Mosby-Year Book, with permission.)

23.5 Synchronization

Most defibrillators for transchest use have the feature of synchronization, which is an electronic sensing and triggering mechanism for application of the shock during the QRS complex of the EGG. This is required when treating arrhythmias other than ventricular fibrillation, because inadvertent application of a shock during the *T* wave of the EGG often produces ventricular fibrillation. Selection by the operator of the synchronized mode of defibrillator operation will cause the defibrillator to automatically sense the QRS complex and apply the shock during the QRS complex. Furthermore, on the EGG

display, the timing of the shock on the QRS is graphically displayed so the operator can be certain that the shock will not fall during the *T* wave (see Fig. 23.9).

23.6 Automatic External Defibrillators

Automatic external defibrillators (AEDs) are defibrillators that automatically or semiautomatically recognize and treat rapid arrhythmias, usually under emergency conditions. Their operation requires less training than operation of manual defibrillators because the operator need not know which EGG waveforms indicate rhythms requiring a shock. The operator applies adhesive electrodes from the AED to the patient and turns on the AED, which monitors the EGG and determines by built-in signal processing whether or not and when to shock the patient. In a completely automatic mode, the AED does not have a manual control as shown in Fig. 23.3 but instead has an automatic control. In semiautomatic mode, the operator must confirm the shock advisory from the AED to deliver the shock. AEDs have substantial potential for improving the chances of survival from cardiac arrest because they enable emergency personnel, who typically reach the patient before paramedics do, to deliver defibrillating shocks. Furthermore, the reduced training requirements make feasible the operation of AEDs in the home by a family member of a patient at high risk of ventricular fibrillation.

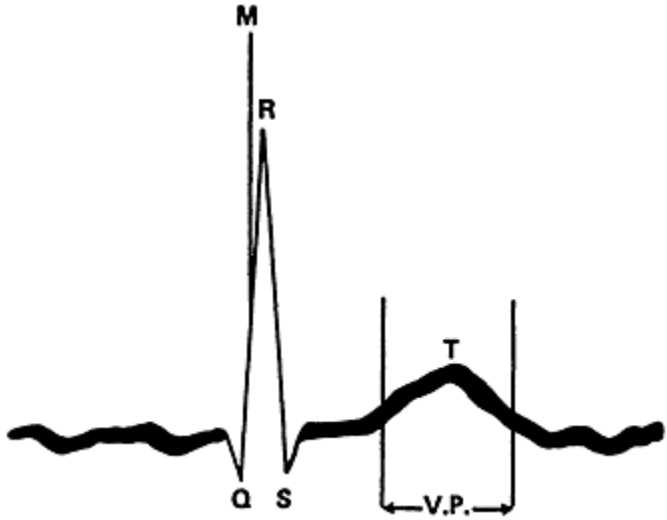


FIGURE 23.9 Timing mark *(M)* as shown on a synchronized defibrillator monitor. The M designates when in the cardiac cycle a shock will be applied. The *T* wave must be avoided, since a shock during the vulnerable period

(V.P.) may fibrillate the ventricles. This tracing shows atrial fibrillation as identified by the irregular wavy baseline of the EGG. (Modified from Feinberg B. 1980. *Handbook Series in Clinical Laboratory Science,* Vol 2, Boca Raton, FL, CRC Press, with permission.)

23.7 Defibrillator Safety

Defibrillators are potentially dangerous devices because of their high electrical output characteristics. The danger to the patient of unsynchronized shocks has already been presented, as has the synchronization design to prevent inadvertent precipitation of fibrillation by a cardioversion shock applied during the *T* wave.

There are other safety issues. Improper technique may result in accidental shocking of the operator or other personnel in the vicinity, if someone is in contact with the electrical discharge pathway. This may occur if the operator is careless in holding the discharge electrodes or if someone is in contact with the patient or with a metal bed occupied by the subject when the shock is applied. Proper training and technique is necessary to avoid this risk.

Another safety issue is that of producing damage to the patient by application of excessively strong or excessively numerous shocks. Although cardiac damage has been reported after high-intensity and repetitive shocks to experimental animals and human patients, it is generally held that significant cardiac damage is unlikely if proper clinical procedures and guidelines are followed.

Failure of a defibrillator to operate correctly may also be considered a safety issue, since inability of a defibrillator to deliver a shock in the absence of a replacement unit means loss of the opportunity to resuscitate the patient. A recent review of defibrillator failures found that operator errors, inadequate defibrillator care and maintenance, and, to a lesser extent, component failure accounted for the majority of defibrillator failures [7].

References

1. Tacker WA Jr (ed). 1994. *Defibrillation of the Heart: ICDs, AEDs, and Manual* St. Louis, Mosby-Year Book.
2. Tacker WA Jr, Geddes LA. 1980. *Electrical Defibrillation.* Boca Raton, FL, CRC Press.
3. Emergency Cardiac Care Committees, American Heart Association. 1992. Guidelines for cardiopulmonary resuscitation and emergency cardiac care. *JAMA* 268:2199.
4. American National Standard ANSI/AAMI DF2. 1989 (second edition, revision of ANSI/AAMI DF2–1981). Safety and performance standard: Cardiac defibrillator devices.

5. Canadian National Standard CAN/CSA C22.2 No. 601.2.4-M90. 1990. Medical electrical equipment, part 2: Particular requirements for the safety of cardiac defibrillators and cardiac defibrillator/monitors.
6. International Standard IEC 601–2-4.1983. Medical electrical equipment, part 2: Particular requirements for the safety of cardiac defibrillators and cardiac defibrillator/monitors.
7. Cummins RO, Chesemore K, White RD, and the Defibrillator Working Group. 1990. Defibrillator failures: Causes of problems and recommendations for improvement. *JAMA* 264:1019.

Further Information

Detailed presentation of material on defibrillator waveforms, algorithms for ECG analysis, and automatic defibrillation using AED's, electrodes, design, clinical use, effects of drugs on shock strength required to defibrillate, damage due to defibrillator shocks, and use of defibrillators during openthorax surgical procedures or trans-esophageal defibrillation are beyond the scope of this chapter. Also, the historical aspects of defibrillation are not presented here. For more information, the reader is referred to the publications at the end of this chapter [1–3]. For detailed description of specific defibrillators with comparisons of features, the reader is referred to articles from *Health Devices*, a monthly publication of ECRI, 5200 Butler Pike, Plymouth Meeting, PA. For American, Canadian, and European defibrillator standards, the reader is referred to published standards [3–6] and Charbonnier's discussion of standards [1].

24
Implantable Defibrillators

Edwin G.Duffin
Medtronic, inc.

The implantable cardioverter defibrillator (ICD) is a therapeutic device that can detect ventricular tachycardia or fibrillation and automatically deliver high-voltage (750 V) shocks that will restore normal sinus rhythm. Advanced versions also provide low-voltage (5 to 10 V) pacing stimuli for painless termination of ventricular tachycardia and for management of bradyarrhythmias. The proven efficacy of the automatic implantable defibrillator has placed it in the mainstream of therapies for the prevention of sudden arrhythmic cardiac death.

The implantable defibrillator has evolved significantly since first appearing in 1980. The newest devices can be implanted in the patient's pectoral region and use electrodes that can be inserted transvenously, eliminating the traumatic thoracotomy required for placement of the earlier epicardial electrode systems. Transvenous systems provide rapid, minimally invasive implants with high assurance of success and greater patient comfort. Advanced arrhythmia detection algorithms offer a high degree of sensitivity with reasonable specificity, and extensive monitoring is provided to document performance and to facilitate appropriate programming of arrhythmia detection and therapy parameters. Generator longevity can now exceed 4 years, and the cost of providing this therapy is declining.

24.1 Pulse Generators

The implantable defibrillator consists of a primary battery, high-voltage capacitor bank, and sensing and control circuitry housed in a hermetically sealed titanium case. Commercially available devices weigh between 197 and 237 grams and range in volume from 113 to 145 cm^3. Clinical trials are in progress on devices with volumes ranging from 178 cm^3 to 60 cm^3 and weights between 275 and 104 grams. Further size reductions will be achieved with the introduction of improved capacitor and integrated circuit technologies and lead systems offering lower pacing and defibrillation thresholds. Progress should parallel that made with antibradycardia pacemakers that have evolved from 250-gram, nonprogrammable, VOO units with 600-μJ pacing outputs to 26-gram, multiprogrammable, DDDR units with dual 25-μJ outputs.

Implantable defibrillator circuitry must include an amplifier, to allow detection of the millivolt-range cardiac electrogram signals; noninvasively programmable processing and control functions, to evaluate the sensed cardiac activity and to direct generation and

0-8493-1813-0/03/$0.00+$1.50

delivery of the therapeutic energy; high-voltage switching capability; dc-dc conversion functions to step up the low battery voltages; random access memories, to store appropriate patient and device data; and radiofrequency telemetry systems, to allow communication to and from the implanted device. Monolithic integrated circuits on hybridized substrates have made it possible to accomplish these diverse functions in a commercially acceptable and highly reliable form.

Defibrillators must convert battery voltages of approximately 6.5 V to the 600 to 750 V needed to defibrillate the heart. Since the conversion process cannot directly supply this high voltage at current strengths needed for defibrillation, charge is accumulated in relatively large (≈85 to 120 μF effective capacitance) aluminum electrolytic capacitors that account for 20 to 30% of the volume of a typical defibrillator. These capacitors must be charged periodically to prevent their dielectric from deteriorating. If this is not done, the capacitors become electrically leaky, yielding excessively long charge times and delay of therapy. Early defibrillators required that the patient return to the clinic periodically to have the capacitors reformed, whereas newer devices do this automatically at preset or programmable times. Improved capacitor technology, perhaps ceramic or thin-film, will eventually offer higher storage densities, greater shape variability for denser component packaging, and freedom from the need to waste battery capacity performing periodic reforming charges. Packaging density has already improved from 0.03 J/cm^3 for devices such as the early cardioverter to 0.43 J/cm^3 with some investigational ICDs. Capacitors that allow conformal shaping could readily increase this density to more than 0.6 J/cm^3.

Power sources used in defibrillators must have sufficient capacity to provide 50 to 400 full-energy charges (≈34 J) and 3 to 5 years of bradycardia pacing and background circuit operation. They must have a very low internal resistance in order to supply the relatively high currents needed to charge the defibrillation capacitors in 5 to 15 s. This generally requires that the batteries have large surface area electrodes and use chemistries that exhibit higher rates of internal discharge than those seen with the lithium iodide batteries used in pacemakers. The most commonly used defibrillator battery chemistry is lithium silver vanadium oxide.

24.2 Electrode Systems (“Leads”)

Early implantable defibrillators utilized patch electrodes (typically a titanium mesh electrode) placed on the surface of the heart, requiring entry through the chest (Fig. 24.1). This procedure is associated with approximately 3 to 4% perioperative mortality, significant hospitalization time and complications, patient discomfort, and high costs. Although subcostal, subxiphoid, and thoracoscopic techniques can minimize the surgical procedure, the ultimate solution has been development of fully transvenous lead systems with acceptable defibrillation thresholds.

Currently available transvenous leads are constructed much like pacemaker leads, using polyurethane or silicone insulation and platinum-iridium electrode materials. Acceptable thresholds are obtained in 67 to 95% of patients, with mean defibrillation thresholds ranging from 10.9 to 18.1 J. These lead systems use a combination of two or more electrodes located in the right ventricular apex, the superior vena cava, the coronary

sinus, and sometimes, a subcutaneous patch electrode is placed in the chest region. These leads offer advantages beyond the avoidance of major surgery. They are easier to remove should there be infections or a need for lead system revision. The pacing thresholds of current transvenous defibrillation electrodes are typically 0.96±0.39 V, and the electrogram amplitudes are on the order of 16.4±6.4 mV. The eventual application of steroid-eluting materials in the leads should provide increased pacing efficiency with transvenous lead systems, thereby reducing the current drain associated with pacing and extending pulse generator longevity.

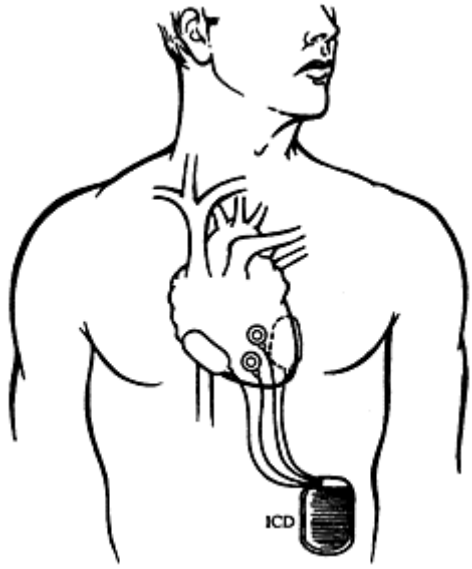


FIGURE 24.1 Epicardial ICD systems typically use two or three large defibrillating patch electrodes placed on the epicardium of the left and right ventricles and a pair of myocardial electrodes for detection and pacing. The generator is usually placed in the abdomen. (Copyright Medtronic, Inc. Used with permission.)

Lead systems are being refined to simplify the implant procedures. One approach is the use of a single catheter having a single right ventricular low-voltage electrode for pacing and detection, and a pair of high-voltage defibrillation electrodes spaced for replacement in the right ventricle and in the superior vena cava (Fig. 24.2*a*). A more recent approach parallels that used for unipolar pacemakers. A single right-ventricular catheter having bipolar pace/sense electrodes and one right ventricular high-voltage electrode is used in conjunction with a defibrillator housing that serves as the second

high-voltage electrode (Fig. 24.2*b*). Mean biphasic pulse defibrillation thresholds with the generator-electrode placed in the patient's left pectoral region are reported to be 9.8±6.6 J (*n*=102). This approach appears to be practicable only with generators suitable for pectoral placement, but such devices will become increasingly available.

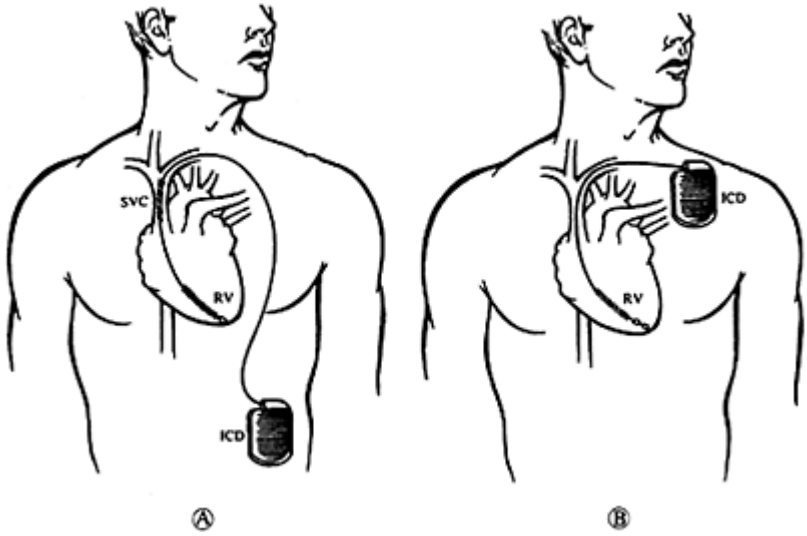


FIGURE 24.2 The latest transvenous fibrillation systems employ a single catheter placed in the right ventricular apex. In panel *a*, a single transvenous catheter provides defibrillation electrodes in the superior vena cava and in the right ventricle. This catheter provides a single pace/sense electrode that is used in conjunction with the right ventricular high-voltage defibrillation electrode for arrhythmia detection and antibradycardia/antitachycardia pacing (a configuration that is sometimes referred to as *integrated bipolar*). With pulse generators small enough to be placed in the pectoral region, defibrillation can be achieved by delivering energy between the

generator housing and one high-voltage electrode in the right ventricle (analogous to unipolar pacing) as is shown in panel *b*. This catheter provided bipolar pace/sense electrodes for arrhythmia detection and antibradycardia/antitachycardia pacing. (Copyright Medtronic, Inc. Used with permission.)

24.3 Arrhythmia Detection

Most defibrillator detection algorithms rely primarily on heart rate to indicate the presence of a treatable rhythm. Additional refinements sometimes include simple morphology assessments, as with the probability density function, and analysis of rhythm stability and rate of change in rate.

The probability density function evaluates the percentage of time that the filtered ventricular electrogram spends in a window centered on the baseline. The rate-of-change-in-rate or *onset* evaluation discriminates sinus tachycardia from ventricular tachycardia on the basis of the typically gradual acceleration of sinus rhythms versus the relatively abrupt acceleration of many pathologic tachycardias. The *rate stability* function is designed to bar detection of tachyarrhythmias as long as the variation in ventricular rate exceeds a physician-programmed tolerance, thereby reducing the likelihood of inappropriate therapy delivery in response to atrial fibrillation. This concept appears to be one of the more successful detection algorithm enhancements.

Because these additions to the detection algorithm reduce sensitivity, some defibrillator designs offer a supplementary detection mode that will trigger therapy in response to any elevated ventricular rate of prolonged duration. These *extended-high-rate* algorithms bypass all or portions of the normal detection screening, resulting in low specificity for rhythms with prolonged elevated rates such as exercise-induced sinus tachycardia. Consequently, use of such algorithms generally increases the incidence of inappropriate therapies.

Improvements in arrhythmia detection specificity are desirable, but they must not decrease the excellent sensitivity offered by current algorithms. The anticipated introduction of defibrillators incorporating dual-chamber pacemaker capability will certainly help in this quest, since it will then be possible to use atrial electrograms in the rhythm classification process. It would also be desirable to have a means of evaluating the patient's hemodynamic tolerance of the rhythm, so that the more comfortable pacing sequences could be used as long as the patient was not syncopal, yet branch quickly to a definitive shock should the patient begin to lose consciousness.

Although various enhanced detection processes have been proposed, many have not been tested clinically, in some cases because sufficient processing power was not available in implantable systems, and in some cases because sensor technology was not

yet ready for chronic implantation. Advances in technology may eventually make some of these very elegant proposals practicable. Examples of proposed detection enhancements include extended analyses of cardiac event timing (PR and RR stability, AV interval variation, temporal distribution of atrial electrogram intervals and of ventricular electrogram intervals, timing differences and/ or coherency of multiple ventricular electrograms, ventricular response to a provocative atrial extrastimuli), electrogram waveform analyses (paced depolarization integral, morphology analyses of right ventricular or atrial electrograms), analyses of hemodynamic parameters (right ventricular pulsatile pressure, mean right atrial and mean right ventricular pressures, wedge coronary sinus pressure, static right ventricular pressure, right atrial pressure, right ventricular stroke volume, mixed venous oxygen saturation and mixed venous blood temperature, left ventricular impedance, intramyocardial pressure gradient, aortic and pulmonary artery flow), and detection of physical motion.

Because defibrillator designs are intentionally biased to overtreat in preference to the life-threatening consequences associated with failure to treat, there is some incidence of inappropriate therapy delivery. Unwarranted therapies are usually triggered by supraventricular tachyarrhythmias, especially atrial fibrillation, or sinus tachycardia associated with rates faster than the ventricular tachycardia detection rate threshold. Additional causes include nonsustained ventricular tachycardia, oversensing of *T* waves, double counting of *R* waves and pacing stimuli from brady pacemakers, and technical faults such as loose lead-generator connections or lead fractures.

Despite the bias for high detection sensitivity, undersensing does occur. It has been shown to result from inappropriate detection algorithm programming, such as an excessively high tachycardia detection rate; inappropriate amplifier gain characteristics; and electrode designs that place the sensing terminals too close to the high-voltage electrodes with a consequent reduction in electrogram amplitude following shocks. Undersensing can also result in the induction of tachycardia should the amplifier gain control algorithm result in undersensing of sinus rhythms.

24.4 Arrhythmia Therapy

Pioneering implantable defibrillators were capable only of defibrillation shocks. Subsequently, synchronized cardioversion capability was added. Antibradycardia pacing had to be provided by implantation of a standard pacemaker in addition to the defibrillator, and, if antitachycardia pacing was prescribed, it was necessary to use an antitachycardia pacemaker. Several currently marketed implantable defibrillators offer integrated ventricular demand pacemaker function and tiered antiarrhythmia therapy (pacing/ cardioversion/defibrillation). Various burst and ramp antitachycardia pacing algorithms are offered, and they all seem to offer comparably high success rates. These expanded therapeutic capabilities improve patient comfort by reducing the incidence of shocks in conscious patients, eliminate the problems and discomfort associated with implantation of multiple devices, and contribute to a greater degree of success, since the prescribed regimens can be carefully tailored to specific patient needs. Availability of devices with antitachy pacing capability significantly increases the acceptability of the implantable defibrillator for patients with ventricular tachycardia.

Human clinical trials have shown that biphasic defibrillation waveforms are more effective than monophasic waveforms, and newer devices now incorporate this characteristic. Speculative explanations for biphasic superiority include the large voltage change at the transition from the first to the second phase or hyperpolarization of tissue and reactivation of sodium channels during the initial phase, with resultant tissue conditioning that allows the second phase to more readily excite the myocardium.

Antitachycardia pacing and cardioversion are not uniformly successful. There is some incidence of ventricular arrhythmia acceleration with antitachycardia pacing and cardioversion, and it is also not unusual for cardioversion to induce atrial fibrillation that in turn triggers unwarranted therapies. An ideal therapeutic solution would be one capable of preventing the occurrence of tachycardia altogether. Prevention techniques have been investigated, among them the use of precisely timed subthreshold stimuli, simultaneous stimulation at multiple sites, and pacing with elevated energies at the site of the tachycardia, but none has yet proven practical.

The rudimentary VVI antibradycardia pacing provided by current defibrillators lacks rate responsiveness and atrial pacing capability. Consequently, some defibrillator patients require implantation of a separate dual-chamber pacemaker for hemodynamic support. It is inevitable that future generations of defibrillators will offer dual-chamber pacing capabilities.

Atrial fibrillation, occurring either as a consequence of defibrillator operation or as a natural progression in many defibrillator patients, is a major therapeutic challenge. It is certainly possible to adapt implantable defibrillator technology to treat atrial fibrillation, but the challenge is to do so without causing the patient undue discomfort. Biphasic waveform defibrillation of acutely induced atrial fibrillation has been demonstrated in humans with an 80% success rate at 0.4 J using *epicardial* electrodes. Stand-alone atrial defibrillators are in development, and, if they are successful, it is likely that this capability would be integrated into the mainstream ventricular defibrillators as well. However, most conscious patients find shocks above 0.5 J to be very unpleasant, and it remains to be demonstrated that a clinically acceptable energy level will be efficacious when applied with transvenous electrode systems to spontaneously occurring atrial fibrillation. Moreover, a stand-alone atrial defibrillator either must deliver an atrial shock with complete assurance of appropriate synchronization to ventricular activity or must restrict the therapeutic energy delivery to atrial structures in order to prevent inadvertent induction of a malignant ventricular arrhythmia.

24.5 Implantable Monitoring

Until recently, defibrillator data recording capabilities were quite limited, making it difficult to verify the adequacy of arrhythmia detection and therapy settings. The latest devices record electrograms and diagnostic channel data showing device behavior during multiple tachyarrhythmia episodes. These devices also include counters (number of events detected, success and failure of each programmed therapy, and so on) that present a broad, though less specific, overview of device behavior (Fig. 24.3). Monitoring

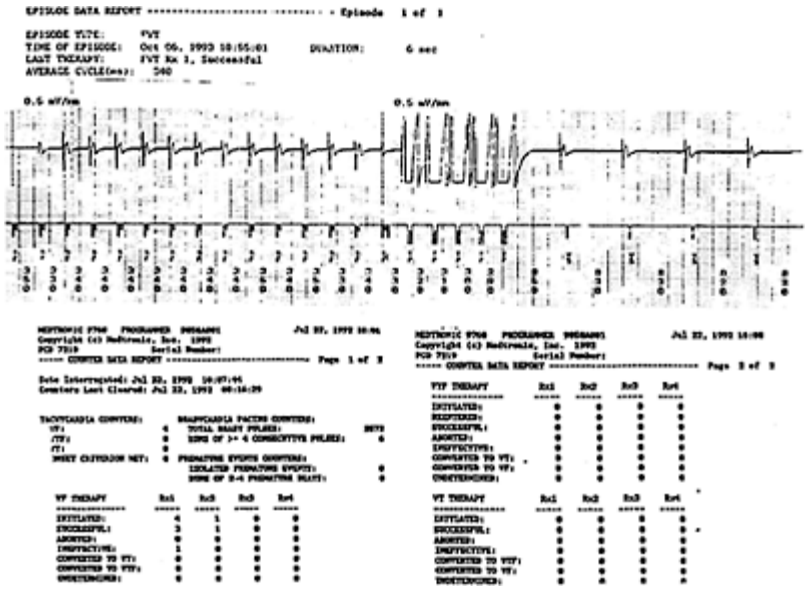


FIGURE 24.3 Typical data recorded by an implantable defibrillator include stored intracardiac electrograms with annotated markers indicating cardiac intervals, paced and sensed events, and device classification of events (TF= fast tachycardia; TP=antitachy pacing stimulus; VS=sensed nontachy ventricular event). In the example, five rapid pacing pulses convert a ventricular tachycardia with a cycle length of 340 ms into sinus rhythm with a cycle length of 830 ms. In the lower portion of the figure is an example of the summary data collected by the ICD, showing detailed counts of the performance of the various therapies (Rx) for ventricular tachycardia (VT), fast ventricular (VTF), and ventricular (VF). (Copyright Medtronic, Inc. Used with permission.)

capability in some of the newest devices appears to be the equivalent of 32 Kbytes of random access memory, allowing electrogram waveform records of approximately 2-min duration, with some opportunity for later expansion by judicious selection of sampling rates and data compression techniques. Electrogram storage has proven useful for documenting false therapy delivery due to atrial fibrillation, lead fractures, and sinus tachycardia, determining the triggers of arrhythmias; documenting rhythm accelerations in response to therapies; and demonstrating appropriate device behavior when treating asymptomatic rhythms.

Electrograms provide useful information by themselves, yet they cannot indicate how the device interpreted cardiac activity. Increasingly, electrogram records are being supplemented with event markers that indicate how the device is responding on a beat-by-beat basis. These records can include measurements of the sensed and paced intervals, indication as to the specific detection zone an event falls in, indication of charge initiation, and other device performance data.

24.6 Follow-up

Defibrillator patients and their devices require careful follow-up. In one study of 241 ICD patients with epicardial lead systems, 53% of the patients experienced one or more complications during an average exposure of 24 months. These complications included infection requiring device removal in 5%, postoperative respiratory complications in 11%, postoperative bleeding and/or thrombosis in 4%, lead system migration or disruption in 8%, and documented inappropriate therapy delivery, most commonly due to atrial fibrillation, in 22%. A shorter study of 80 patients with transvenous defibrillator systems reported no postoperative pulmonary complications, transient nerve injury (1%), asymptomatic subclavian vein occlusion (2.5%), pericardial effusion (1%), subcutaneous patch pocket hematoma (5%), pulse generator pocket infection (1%), lead fracture (1%), and lead system dislodgement (10%). During a mean follow-up period of 11 months, 7.5% of the patients in this series experienced inappropriate therapy delivery, half for atrial fibrillation and the rest for sinus tachycardia.

Although routine follow-up can be accomplished in the clinic, detection and analysis of transient events depends on the recording capabilities available in the devices or on the use of various external monitoring equipment.

24.7 Economics

The annual cost of ICD therapy is dropping as a consequence of better longevity and simpler implantation techniques. Early generators that lacked programmability, antibradycardia pacing capability, and event recording had 62% survival at 18 months and 2% at 30 months. Some recent programmable designs that include VVI pacing capability and considerable event storage exhibit 96.8% survival at 48 months. It has been estimated that an increase in generator longevity from 2 to 5 years would lower the cost per life-year saved by 55% in a hypothetical patient population with a 3-year sudden

mortality of 28%. More efficient energy conversion circuits and finer line-width integrated circuit technology with smaller, more highly integrated circuits and reduced current drains will yield longer-lasting defibrillators while continuing the evolution to smaller volumes.

Cost of the implantation procedure is clearly declining as transvenous lead systems become commonplace. Total hospitalization duration, complication rates, and use of costly hospital operating rooms and intensive-care facilities all are reduced, providing significant financial benefits. One study reported requiring half the in tensive-care unit time and a reduction in total hospitalization from 26 to 15 days when comparing transvenous to epicardial approaches. Another center reported a mean hospitalization stay of 6 days for patients receiving transvenous defibrillation systems.

Increasing sophistication of the implantable defibrillators paradoxically contributes to cost efficacy. Incorporation of single-chamber brady pacing capability eliminates the cost of a separate pacemaker and lead for those patients who need one. Eventually even dual-chamber pacing capability will be available. Programmable detection and therapy features obviate the need for device replacement that was required when fixed parameter devices proved to be inappropriately specified or too inflexible to adapt to a patient's physiologic changes.

Significant cost savings may be obtained by better patient selection criteria and processes, obviating the need for extensive hospitalization and costly electrophysiologic studies prior to device implantation in some patient groups. One frequently discussed issue is the prophylactic role that implantable defibrillators will or should play. Unless a means is found to build far less expensive devices that can be placed with minimal time and facilities, the life-saving yield for prophylactic defibrillators will have to be high if they are to be cost-effective. This remains an open issue.

24.8 Conclusion

The implantable defibrillator is now an established and powerful therapeutic tool. The transition to pectoral implants with biphasic waveforms and efficient yet simple transvenous lead systems is simplifying the implant procedure and drastically reducing the number of unpleasant VF inductions required to demonstrate adequate system performance. These advances are making the implantable defibrillator easier to use, less costly, and more acceptable to patients and their physicians.

Acknowledgment

Portions of this text are derived from Duffin EG, Barold SS. 1994, Implantable cardioverter-defibrillators: An overview and future directions, Chapter 28 of Singer I (ed), *Implantable Cardioverter-Defibrillator*, and are used with permission of Futura Publishing Company, Inc.

References

Josephson M, Wellens H (eds). 1992. *Tachycardias: Mechanisms and Management.* Mount Kisco, NY, Futura Publishing.

Kappenberger L, Lindemans F (eds). 1992. *Practical Aspects of Staged Therapy Defibrillators.* Mount Kisco, NY, Futura Publishing.

Singer I (ed). 1994. *Implantable Cardioverter-Defibrillator.* Mount Kisco, NY, Futura Publishing,

lacker W (ed). 1994. *Defibrillation of the Heart: ICD's, AED's, and Manual.* St. Louis, Mosby.

PACE, 14:865. (Memorial issue on implantable defibrillators honoring Michel Mirowski.)

25
Electrosurgical Devices

Jeffrey L.Eggleston
Valley lab, Inc.

Wolf W.von Maltzahn
Whitaker Foundation

An electrosurgical unit (ESU) passes high-frequency electric currents through biologic tissues to achieve specific surgical effects such as cutting, coagulation, or desiccation. Although it is not completely understood how electrosurgery works, it has been used since the 1920s to cut tissue effectively while at the same time controlling the amount of bleeding. Cutting is achieved primarily with a continuous sinusoidal waveform, whereas coagulation is achieved primarily with a series of sinusoidal wave packets. The surgeon selects either one of these waveforms or a blend of them to suit the surgical needs. An electrosurgical unit can be operated in two modes, the monopolar mode and the bipolar mode. The most noticeable difference between these two modes is the method in which the electric current enters and leaves the tissue. In the monopolar mode, the current flows from a small active electrode into the surgical site, spreads through the body, and returns to a large dispersive electrode on the skin. The high-current density in the vicinity of the active electrode achieves tissue cutting or coagulation, whereas the low current density under the dispersive electrode causes no tissue damage. In the bipolar mode, the current flows only through the tissue held between two forceps electrodes. The monopolar mode is used for both cutting and coagulation. The bipolar mode is used primarily for coagulation.

This chapter beings with the theory of operation for electrosurgical units, outlines various modes of operation, and gives basic design details for electronic circuits and electrodes. It then describes how improper application of electrosurgical units can lead to hazardous situations for both the operator and the patient and how such hazardous situations can be avoided or reduced through proper monitoring methods. Finally, the chapter gives an update on current and future developments and applications.

25.1 Theory of Operation

In principle, electrosurgery is based on the rapid heating of tissue. To better understand the thermody-namic events during electrosurgery, it helps to know the general effects of heat on biologic tissue. Consider a tissue volume that experiences a temperature increase from normal body temperature to 45°C within a few seconds. Although the cells in this

0-8493-1813-0/03/$0.00+$ 1.50

tissue volume show neither microscopic nor macroscopic changes, some cytochemical changes do in fact occur. However, these changes are reversible, and the cells return to their normal function when the temperature returns to normal values. Above 45°C, irreversible changes take place that inhibit normal cell functions and lead to cell death. First, between 45°C and 60°C, the proteins in the cell lose their quaternary configuration and solidify into a glutinous substance that resembles the white of a hard-boiled egg. This process, termed *coagulation*, is accompanied by tissue blanching. Further increasing the temperature up to 100°C leads to tissue drying; that is, the aqueous cell contents evaporate. This process is called *desiccation*. If the temperature is increased beyond 100°C, the solid contents of the tissue reduce to carbon, a process referred to as *carbonization*. Tissue damage depends not only on temperature, however, but also on the length of exposure to heat. Thus, the overall temperature-induced tissue damage is an integrative effect between temperature and time that is expressed mathematically by the Arrhenius relationship, where an exponential function of temperature is integrated over time [1].

In the monopolar mode, the active electrode either touches the tissue directly or is held a few millimeters above the tissue. When the electrode is held above the tissue, the electric current bridges the air gap by creating an electric discharge arc. A visible arc forms when the electric field strength exceeds 1 kV/mm in the gap and disappears when the field strength drops below a certain threshold level.

When the active electrode touches the tissue and the current flows directly from the electrode into the tissue without forming an arc, the rise in tissue temperature follows the bioheat equation,

$$T - T_o = \frac{1}{\sigma \rho c} J^2 t \,, \tag{25.1}$$

where T and T_o are the final and initial temperatures (K), σ is the electrical conductivity (S/m), ρ is the tissue density (kg/m^3), c is the specific heat of the tissue (Jkg^{-1}K^{-1}), J is the current density (A/m^2), and t is the duration of heat applications [1]. The bioheat equation is valid for short application times where secondary effects such as heat transfer to surrounding tissues, blood perfusion, and metabolic heat can be neglected. According to Eq. (25.1), the surgeon has primarily three means of controlling the cutting or coagulation effect during electrosurgery: the contact area between active electrode and tissue, the electrical current density, and the activation time. In most commercially available electrosurgical generators, the output variable that can be adjusted is power. This power setting, in conjunction with the output power vs. tissue impedance characteristics of the generator, allow the surgeon some control over current. Table 25.1 lists typical output power and mode settings for various surgical procedures. Table 25.2 lists some typical impedance ranges seen during use of an ESU in surgery. The values are shown as ranges because the impedance increases as the tissue dries out, and at the same time, the output power of the ESU decreases. The surgeon may control current density by selection of the active electrode type and size.

25.2 Monopolar Mode

A continuous sinusoidal waveform cuts tissue with very little hemostasis. This waveform is simply called *cut* or *pure cut*. During each positive and negative swing of the sinusoidal waveform, a new discharge arc forms and disappears at essentially the same tissue location. The electric current concentrates at this tissue location, causing a sudden increase in temperature due to resistive heating. The rapid rise in temperature then vaporizes intracellular fluids, increases cell pressure, and ruptures the cell membrane, thereby parting the tissue. This chain of events is confined to the vicinity of the arc, because from there the electric current spreads to a much larger tissue volume, and the current density is no longer high enough to cause resistive heating damage. Typical output values for ESUs, in cut and other modes, are shown in Table 25.3.

Experimental observations have shown that more hemostasis is achieved when cutting with an interrupted sinusoidal waveform or amplitude modulated continuous waveform. These waveforms are typically called *blend* or *blended cut*. Some ESUs offer a choice of blend waveforms to allow the surgeon to select the degree of hemostasis desired.

When a continuous or interrupted waveform is used in contact with the tissue and the output voltage current density is too low to sustain arcing, desiccation of the tissue will occur. Some ESUs have a distinct mode for this purpose called *desiccation* or *contact coagulation*.

TABLE 25.1 Typical ESU Power Settings for Various Surgical Procedures

Power-Level Range	Procedures
Low power	
<30 W cut	Neurosurgery
<30 W coag	Dermatology
	Plastic surgery
	Oral surgery
	Laparoscopic sterilization
	Vasectomy
Medium power	
30 W-150 W cut	General surgery
30 W-70 W coag	Laparotomies
	Head and neck surgery (ENT)
	Major orthopedic surgery
	Major vascular surgery
	Routine thoracic surgery
	Polypectomy

High power	
>150 W cut	Transurethral resection procedures (TURPs)
>70 W coag	Thoracotomies
	Ablative cancer surgery
	Mastectomies

Note: Ranges assume the use of a standard blade electrode. Use of a needle electrode, or other small cur rent-concentrating electrode, allows lower settings to be used; users are urged to use the lowest setting that provides the desired clinical results.

TABLE 25.2 Typical Impedance Ranges Seen During Use of an ESU in Surgery

Cut Mode Application	Impedance Range (Ω)
Prostate tissue	400–1700
Oral cavity	1000–2000
Liver tissue	
Muscle tissue	
Gall bladder	1500–2400
Skin tissue	1700–2500
Bowel tissue	2500–3000
Periosteum	
Mesentery	3000–4200
Omentum	
Adipose tissue	3500–4500
Scar tissue	
Adhesions	
Coag Mode Application	
Contact coagulation to stop bleeding	100–1000

In noncontact coagulation, the duty cycle of an interrupted waveform and the crest factor (ratio of peak voltage to rms voltage) influence the degree of hemostasis. While a continuous waveform reestablishes the arc at essentially the same tissue location concentrating the heat there, an interrupted waveform causes the arc to reestablish itself at different tissue locations. The arc seems to dance from one location to the other raising the temperature of the top tissue layer to coagulation levels. These waveforms are called *fulguration* or *spray*. Since the current inside the tissue spreads very quickly from the point where the arc strikes, the heat concentrates in the top layer, primarily desiccating tissue and causing some

TABLE 25.3 Typical Output Characteristics of ESUs

	Output Voltage Range Open Circuit, $V_{peak\text{-}peak}$, V	Output Power Range, W	Frequency, kHz	Crest Factor $\left(\frac{V_{peak}}{V_{rms}}\right)$	Duty Cycle
Monopolar modes					
Cut	200–5000	1–400	300–1750	1.4–2.1	100%
Blend	1500–5800	1–300	300–1750	2.1–6.0	25–80%
Desiccate	400–6500	1–200	240–800	3.5–6.0	50–100%
Fulgurate/spray	6000–12000	1–200	300–800	6.0–20.0	10–70%
Bipolar mode					
Coagulate/desiccate	200–1000	1–70	300–1050	1.6–12.0	25–100%

carbonization. During surgery, a surgeon can easily choose between cutting, coagulation, or a combination of the two by activating a switch on the grip of the active electrode or by use of a footswitch.

25.3 Bipolar Mode

The bipolar mode concentrates the current flow between the two electrodes, requiring considerably less power for achieving the same coagulation effect than the monopolar mode. For example, consider coagulating a small blood vessel with 3-mm external diameter and 2-mm internal diameter, a tissue resistivity of 360 Ωcm, a contract area of 2×4 mm, and a distance between the forceps tips of 1 mm. The tissue resistance between the forceps is 450 Ω as calculated from $R=\rho L/A$, where ρ is the resistivity, L is the distance between the forceps, and A is the contact area. Assuming a typical current density of 200 mA/cm^2, then a small current of 16 mA, a voltage of 7.2 V, and a power level of 0.12 W suffice to coagulate this small blood vessel. In contrast, during monopolar coagulation, current levels of 200 mA and power levels of 100 W or more are not uncommon to achieve the same surgical effect. The temperature increase in the vessel tissue follows the bioheat equation, Eq. (25.1). If the specific heat of the vessel tissue is 4.2 Jg^{-1}k^{-1} and the tissue density is 1 g/cm^3, then the temperature of the tissue between the forceps increases from 37°C to 57°C in 5.83 s. When the active electrode touches the tissue, less tissue damage occurs during coagulation, because the charring and carbonization that accompanies fulguration is avoided.

25.4 ESU Design

Modern ESUs contain building blocks that are also found in other medical devices, such as microprocessors, power supplies, enclosures, cables, indicators, displays, and alarms. The main building blocks unique to ESUs are control input switches, the high-frequency power amplifier, and the safety monitor. The first two will be discussed briefly here, and the latter will be discussed later.

Control input switches include front-panel controls, footswitch controls, and handswitch controls. In order to make operating an ESU more uniform between models and manufacturers, and to reduce the possibility of operator error, the ANSI/AAMI HF-18 standard [5] makes specific recommendations concerning the physical construction and location of these switches and prescribes mechanical and electrical performance standards. For instance, front-panel controls need to have their function identified by a permanent label and their output indicated on alphanumeric displays or on graduated scales; the pedals of foot switches need to be labeled and respond to a specified activation force; and if the active electrode handle incorporates two finger switches, their position has to correspond to a specific function. Additional recommendations can be found in Reference [5].

Four basic high-frequency power amplifiers are in use currently; the somewhat dated vacuum tube/ spark gap configuration, the parallel connection of a bank if bipolar power transistors, the hybrid connection of parallel bipolar power transistors cascaded with metal oxide silicon field effect transistors (MOSFETs), and the bridge connection of MOSFETs. Each has unique properties and represents a stage in the evolution of ESUs.

In a vacuum tube/spark gap device, a tuned-plate, tuned-grid vacuum tube oscillator is used to generate a continuous waveform for use in cutting. This signal is introduced to the patient by an adjustable isolation transformer. To generate a waveform for fulguration, the power-supply voltage is elevated by a step-up transformer to about 1600 V rms, which then connects to a series of spark gaps. The voltage across the spark gaps is capacitively coupled to the primary of an isolation transformer. The RLC circuit created by this arrangement generates a high crest factor, damped sinusoidal, interrupted waveform. One can adjust the output power and characteristics by changing the turns ratio or tap on the primary and/ or secondary side of the isolation transformer, or by changing the spark gap distance.

In those devices that use a parallel bank of bipolar power transistors, the transistors are arranged in a Class A configuration. The bases, collectors, and emitters are all connected in parallel, and the collective base node is driven through a current-limiting resistor. A feedback RC network between the base node and the collector node stabilizes the circuit. The collectors are usually fused individually before the common node connects them to one side of the primary of the step-up transformer. The other side of the primary is connected to the high-voltage power supply. A capacitor and resistor in parallel to the primary create a resonance tank circuit that generates the output waveform at a specific frequency. Additional elements may be switched in and out of the primary parallel RLC to alter the output power and waveform for various electrosurgical modes. Small-value resistors between the emitters and ground improve the current sharing between transistors. This configuration sometimes requires the use of matched sets of high-voltage power transistors.

A similar arrangement exists in amplifiers using parallel bipolar transistors cascaded with a power MOSFET. This arrangement is called a *hybrid cascade amplifier*. In this type of amplifier, the collectors of a group if bipolar transistors are connected, via protection diodes, to one side of the primary of the step-up output transformer. The other side of the primary is connected to the high-voltage power supply. The emitters of two or three bipolar transistors are connected, via current limiting resistors, to the drain of an enhancement mode MOSFET. The source of the MOSFET is connected to ground, and the gate of the MOSFET is connected to a voltage-snubbing network driven by a fixed amplitude pulse created by a high-speed MOS driver circuit. The bases of the bipolar transistors are connected, via current control RC networks, to a common variable base voltage source. Each collector and base is separately fused. In cut modes, the gate drive pulse is a fixed frequency, and the base voltage is varied according to the power setting. In the coagulation modes, the base voltage is fixed and the width of the pulses driving the MOSFET is varied. This changes the conduction time of the amplifier and controls the amount of energy imparted to the output transformer and its load. In the coagulation modes and in high-power cut modes, the bipolar power transistors are saturated, and the voltage across the bipolar/MOSFET combination is low. This translates to high efficiency and low power dissipation.

The most common high-frequency power amplifier in use is a bridge connection of MOSFETs. In this configuration, the drains of a series of power MOSFETs are connected, via protection diodes, to one side of the primary of the step-up output transformer. The drain protection diodes protect the MOSFETs against the negative voltage swings of the transformer primary. The other side of the transformer primary is connected to the high-voltage power supply. The sources of the MOSFETs are connected to ground. The gate of each MOSFET has a resistor connected to ground and one to its driver circuitry. The resistor to ground speeds up the discharge of the gate capacitance when the MOSFET is turned on while the gate series resistor eliminates turn-off oscillations. Various combinations of capacitors and/or LC networks can be switched across the primary of the step-up output transformer to obtain different waveforms. In the cut mode, the output power is controlled by varying the high-voltage power supply voltage. In the coagulation mode, the output power is controlled by varying the on time of the gate drive pulse.

25.5 Active Electrodes

The monopolar active electrode is typically a small flat blade with symmetric leading and trailing edges that is embedded at the tip of an insulated handle. The edges of the blade are shaped to easily initiate discharge arcs and to help the surgeon manipulate the incision; the edges cannot mechanically cut tissue. Since the surgeon holds the handle like a pencil, it is often referred to as the “pencil.” Many pencils contain in their handle one or more switches to control the electrosurgical waveform, primarily to switch between cutting and coagulation. Other active electrodes include needle electrodes, loop electrodes, and ball electrodes. Needle electrodes are used for coagulating small tissue volumes like in neurosurgery or plastic surgery. Loop electrodes are used to resect nodular structures such as polyps or to excise tissue samples for pathologic analysis. An

example would be the LLETZ procedure, where the transition zone of the cervix is excised. Electrosurgery at the tip of an endoscope or laparoscope requires yet another set of active electrodes and specialized training of the surgeon.

25.6 Dispersive Electrodes

The main purpose of the dispersive electrode is to return the high-frequency current to the electrosurgical unit without causing harm to the patient. This is usually achieved by attaching a large electrode to the patient's skin away from the surgical site. The large electrode area and a small contact impedance reduce the current density to levels where tissue heating is minimal. Since the ability of a dispersive electrode to avoid tissue heating and burns is of primary importance, dispersive electrodes are often characterized by their *heating factor*. The heating factor describes the energy dissipated under the dispersive electrode per Ω of impedance and is equal to I^2t, where I is the rms current and t is the time of exposure. During surgery a typical value for the heating factor is 3 A^2s, but factors of up to 9 A^2s may occur during some procedures [2].

Two types of dispersive electrodes are in common use today, the resistive type and the capacitive type. In disposable form, both electrodes have a similar structure and appearance. A thin, rectangular metallic foil has an insulating layer on the outside, connects to a gel-like material on the inside, and may be surrounded by an adhesive foam. In the resistive type, the gel-like material is made of an adhesive conductive gel, whereas in the capacitive type, the gel is an adhesive dielectric nonconductive gel. The adhesive foam and adhesive gel layer ensure that both electrodes maintain good skin contact to the patient, even if the electrode gets stressed mechanically from pulls on the electrode cable. Both types have specific advantages and disadvantages. Electrode failures and subsequent patient injury can be attributed mostly to improper application, electrode dislodgment, and electrode defects rather than to electrode design.

25.7 ESU Hazards

Improper use of electrosurgery may expose both the patient and the surgical staff to a number of hazards. By far the most frequent hazards are electric shock and undesired burns. Less frequent are undesired neuromuscular stimulation, interference with pacemakers or other devices, electrochemical effects from direct currents, implant heating, and gas explosions [1, 3].

Current returns to the ESU through the dispersive electrode. If the contact area of the dispersive electrode is large and the current exposure time short, then the skin temperature under the electrode does not rise above 45°C, which has been shown to be the maximum safe temperature [4]. However, to include a safety margin, the skin temperature should not rise more than 6°C above the normal surface temperature of 29 to 33°C. The current density at any point under the dispersive electrode has to be significantly below the recognized burn threshold of 100 mA/cm^2 for 10 seconds.

To avoid electric shock and burns, the American National Standard for Electrosurgical Devices [5] requires that "any electrosurgical generator that provides for a dispersive electrode and that has a rated output power of greater than 50 W shall have at least one patient circuit safety monitor." The most common safety monitors are the contact quality monitor for the dispersive electrode and the patient circuit monitor. A contact quality monitor consists of a circuit to measure the impedance between the two sides of a split dispersive electrode and the skin. A small high-frequency current flows from one section of the dispersive electrode through the skin to the second section of the dispersive electrode. If the impedance between these two sections exceeds a certain threshold, or changes by a certain percentage, an audible alarm sounds, and the ESU output is disabled.

Patient circuit monitors range from simple to complex. The simple ones monitor electrode cable integrity while the complex ones detect any abnormal condition that could result in electrosurgical current flowing in other than normal pathways. Although the output isolation transformer present in most modern ESUs usually provides adequate patient protection, some potentially hazardous conditions may still arise. If a conductor to the dispersive electrode is broken, undesired arcing between the broken conductor ends may occur, causing fire in the operating room and serious patient injury. Abnormal current pathways may also arise from capacitive coupling between cables, the patient, operators, enclosures, beds, or any other conductive surface or from direct connections to other electrodes connected to the patient. The patient circuit monitoring device should be operated from an isolated power source having a maximum voltage of 12 V rms. The most common device is a cable continuity monitor. Unlike the contact quality monitor, this monitor only checks the continuity of the cable between the ESU and the dispersive electrode and sounds an alarm if the resistance in that conductor is greater than 1 kΩ. Another implementation of a patient circuit monitor measures the voltage between the dispersive electrode connection and ground. A third implementation functions similarly to a ground fault circuit interrupter (GFCI) in that the current in the wire to the active electrode and the current in the wire to the dispersive electrode are measured and compared with each other. If the difference between these currents is greater than a preset threshold, the alarm sounds and the ESU is disconnected.

There are other sources of undesired burns. Active electrodes get hot when they are used. After use, the active electrode should be placed in a protective holster, if available, or on a suitable surface to isolate it from the patient and surgical staff. The correct placement of an active electrode will also prevent the patient and/or surgeon from being burned if an inadvertent activation of the ESU occurs (e.g., someone accidentally stepping on a foot pedal). Some surgeons use a practice called *buzzing the hemostat* in which a small bleeding vessel is grasped with a clamp or hemostat and the active electrode touched to the clamp while activating. Because of the high voltages involved and the stray capacitance to ground, the surgeon's glove may be compromised. If the surgical staff cannot be convinced to eliminate the practice of buzzing hemostats, the probability of burns can be reduced by use of a cut waveform instead of a coagulation waveform (lower voltage), by maximizing contact between the surgeon's hand and the clamp, and by not activating until the active electrode is firmly touching the clamp.

Although it is commonly assumed that neuromuscular stimulation ceases or is insignificant at frequencies above 10 kHz, such stimulation has been observed in

anesthetized patients undergoing certain electrosurgical procedures. This undesirable side effect of electrosurgery is generally attributed to non-linear events during the electric arcing between the active electrode and tissue. These events rectify the high-frequency current leading to both dc and low-frequency current components. These current components can reach magnitudes that stimulate nerve and muscle cells. To minimize the probability of unwanted neuromuscular stimulation, most ESUs incorporate in their output circuit a high-pass filter that suppresses dc and low-frequency current components.

The use of electrosurgery means the presence of electric discharge arcs. This presents a potential fire hazard in an operating room where oxygen and flammable gases may be present. These flammable gases may be introduced by the surgical staff (anesthetics or flammable cleaning solutions), or may be generated within the patients themselves (bowel gases). The use of disposable paper drapes and dry surgical gauze also provides a flammable material that may be ignited by sparking or by contact with a hot active electrode. Therefore, prevention of fires and explosions depends primarily on the prudence and judgment of the ESU operator.

25.8 Recent Developments

Electrosurgery is being enhanced by the addition of a controlled column of argon gas in the path between the active electrode and the tissue. The flow of argon gas assists in clearing the surgical site of fluid and improves visibility. When used in the coagulation mode, the argon gas is turned into a plasma allowing tissue damage and smoke to be reduced, and producing a thinner, more flexible eschar. When used with the cut mode, lower power levels may be used.

Many manufacturers have begun to include sophisticated computer-based systems in their ESUs that not only simplify the use of the device but also increase the safety of patient and operator [7]. For instance, in a so-called soft coagulation mode, a special circuit continuously monitors the current between the active electrode and the tissue and turns the ESU output on only after the active electrode has contacted the tissue. Furthermore, the ESU output is turned off automatically, once the current has reached a certain threshold level that is typical for coagulated and desiccated tissue. This feature is also used in a bipolar mode termed *autobipolar*. Not only does this feature prevent arcing at the beginning of the procedure, but it also keeps the tissue from being heated beyond 70°C. Some devices offer a so-called power-peak-system that delivers a very short power peak at the beginning of electrosurgical cutting to start the cutting arc. Other modern devices use continuous monitoring of current and voltage levels to make automatic power adjustments in order to provide for a smooth cutting action from the beginning of the incision to its end. Some manufacturers are developing waveforms and instruments designed to achieve specific clinical results such as bipolar cutting tissue lesioning, and vessel sealing. With the growth and popularity of laparoscopic procedures, additional electrosurgical instruments and waveforms tailored to this surgical specialty should also be expected.

Increased computing power, more sophisticated evaluation of voltage and current waveforms, and the addition of miniaturized sensors will continue to make ESUs more user-friendly and safer.

Defining Terms

Active electrode: Electrode used for achieving desired surgical effect.
Coagulation: Solidification of proteins accompanied by tissue whitening.
Desiccation: Drying of tissue due to the evaporation of intracellular fluids.
Dispersive electrode: Return electrode at which no electrosurgical effect is intended.
Fulguration: Random discharge of sparks between active electrode and tissue surface in order to achieve coagulation and/or desiccation.
Spray: Another term for **fulguration**. Sometimes this waveform has a higher crest factor than that used for fulguration.

References

1. Pearce John A. 1986. *Electrosurgery*, New York, John Wiley.
2. Gerhard Glen C. 1988. Electrosurgical unit. In JG Webster (ed), *Encyclopedia of Medical Devices and Instrumentation*, Vol 2, pp 1180–1203, New York, John Wiley.
3. Gendron Francis G. 1988. *Unexplained Patient Burns: Investigating latrogenic Injuries*, Brea, Calif, Quest Publishing.
4. Pearce JA, Geddes LA, Van Vleet JF, et al. 1983. Skin burns from electrosurgical current. *Med Instrum* 17(3):225.
5. American National Standard for Electrosurgical Devices. 1994. HF18, American National Standards Institute.
6. LaCourse JR, Miller WT III, Vogt M, et al. 1985. Effect of high frequency current on nerve and muscle tissue. *IEEE Trans Biomed Eng* 32:83.
7. Haag R, Cuschieri A. 1993. Recent advances in high-frequency electrosurgery: Development of automated systems. *J R Coll Surg Ednb* 38:354.

Further Information

American National Standards Institute, 1988. International Standard, Medical Electrical Equipment, *Part 1: General Requirements for Safety*, IEC 601–1, 2d ed, New York.
American National Standards Institute. 1991. International Standard, Medical Electrical Equipment, *Part 2: Particular Requirements for the Safety of High Frequency Surgical Equipment*, IEC 601–2-2, 2d ed, New York.
National Fire Protection Association. 1993. *Standard for Health Care Facilities*, NFPA 99.

26
Mechanical Ventilation

Khosrow Rehhehani
The University of Texas at Arlington
and The University of Texas
Southwestern Medical Center
at Dallas

26.1 Introduction

This chapter presents an overview of the structure and function of mechanical ventilators. Mechanical ventilators, which are often also called respirators, are used to artificially ventilate the lungs of patients who are unable to naturally breathe from the atmosphere. In almost 100 years of development, many mechanical ventilators with different designs have been developed [Mushin et al., 1980; Philbeam, 1998]. The very early devices used bellows that were manually operated to inflate the lungs. Today's respirators employ an array of sophisticated components such as microprocessors, fast response servo valves, and precision transducers to perform the task of ventilating the lungs. The changes in the design of ventilators have come about as the result of improvements in engineering the ventilator components and the advent of new therapy modes by clinicians. A large variety of ventilators are now available for short-term treatment of acute respiratory problems as well as long-term therapy for chronic respiratory conditions.

It is reasonable to broadly classify today's ventilators into two groups. The first and indeed the largest group encompasses the intensive care respirators used primarily in hospitals to support patients following certain surgical procedures or assist patients with acute respiratory disorders. The second group includes less complicated machines that are primarily used at home to treat patients with chronic respiratory disorders.

The level of engineering design and sophistication for the intensive care ventilators is higher than the ventilators used for chronic treatment. However, many of the engineering concepts employed in designing intensive care ventilators can also be applied in the simpler chronic care units. Therefore, this presentation focuses on the design of intensive care ventilators; the terms respirator, mechanical ventilator, or ventilator that will be used from this point on refer to the intensive care unit respirators.

At the beginning, the designers of mechanical ventilators realized that the main task of a respirator was to ventilate the lungs in a manner as close to natural respiration as possible. Since natural inspiration is a result of negative pressure in the pleural cavity generated by distention of the diaphragm, designers initially developed ventilators that

0-8493-1813-0/03/$0.00+$1.50

created the same effect. These ventilators are called *negative-pressure ventilators*. However, more modern ventilators use pressures greater than atmospheric pressures to ventilate the lungs; they are known as *positive-pressure ventilators.*

26.2 Negative-Pressure Ventilators

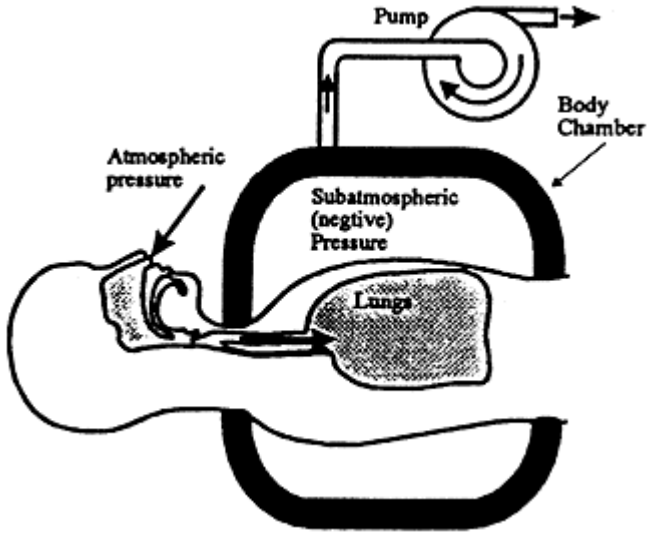


FIGURE 26.1 A simplified illustration of a negative-pressure ventilator.

The principle of operation of a negative-pressure respirator is shown in Fig. 26.1. In this design, the flow of air to the lungs is created by generating a negative pressure around the patient's thoracic cage. The negative pressure moves the thoracic walls outward expanding the intrathoracic volume and dropping the pressure inside the lungs. The pressure gradient between the atmosphere and the lungs causes the flow of atmospheric air into the lungs. The inspiratory and expiratory phases of the respiration are controlled by cycling the pressure inside the body chamber between a subatmospheric level (inspiration) and the atmospheric level (exhalation). Flow of the breath out of the lungs during exhalation is caused by the recoil of thoracic muscles.

Although it may appear that the negative-pressure respirator incorporates the same principles as natural respiration, the engineering implementation of this concept has not been very successful. A major difficulty has been in the design of a chamber for creating negative pressure around the thoracic walls. One approach has been to make the chamber large enough to house the entire body with the exception of the head and neck. Using foam rubber around the patient's neck, an attempt is made to seal the chamber and generate a negative pressure inside the chamber. This design configuration, commonly known as the iron lung, was tried in the 1920s and proved to be deficient in several

aspects. The main drawback was that the negative pressure generated inside the chamber was applied to the chest as well as the abdominal wall, thus creating a venous blood pool in the abdomen and reducing cardiac output.

More recent designs have tried to restrict the application of the negative pressure to the chest walls by designing a chamber that goes only around the chest. However, this has not been successful because obtaining a seal around the chest wall (Fig. 26.1) is difficult.

Negative-pressure ventilators also made the patient less accessible for patient care and monitoring. Further, synchronization of the machine cycle with the patient's effort was has been difficult and they are also typically noisy and bulky [McPherson and Spearman 1990]. These deficiencies of the negative-pressure ventilators have led to the development of the positive-pressure ventilators.

26.3 Positive-Pressure Ventilators

Positive-pressure ventilators generate the inspiratory flow by applying a positive pressure (greater than the atmospheric pressure) to the airways. Figure 26.2 shows a simplified block diagram of a positive-pressure ventilator. During inspiration, the inspiratory flow delivery system creates a positive pressure in the tubes connected to the patient airway, called **patient circuit,** and the exhalation control system closes a valve at the outlet of the tubing to the atmosphere. When the ventilator switches to exhalation, the inspiratory flow delivery system stops the positive pressure and the exhalation system opens the valve to allow the patient's exhaled breath to flow to the atmosphere. The use of a positive pressure gradient in creating the flow allows treatment of patients with high lung resistance and low compliance. As a result, positive-pressure ventilators have been very successful in treating a variety of breathing disorders and have become more popular than negative-pressure ventilators.

Positive-pressure ventilators have been employed to treat patients ranging from neonates to adults. Due to anatomical differences between various patient populations, the ventilators and their modes of

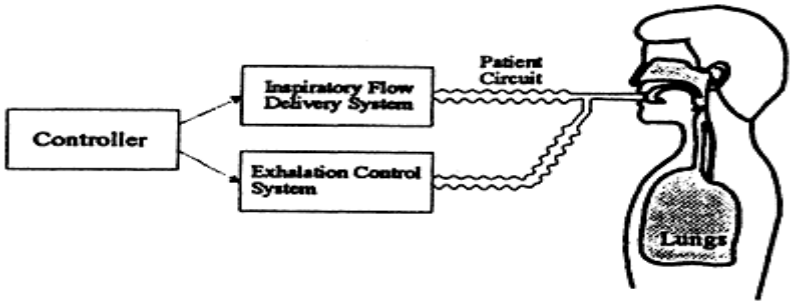


FIGURE 26.2 A simplified diagram of the functional blocks of a positive-pressure ventilator.

treating infants are different than those for adults. Nonetheless, their fundamental design principles are similar and adult ventilators comprise a larger percentage of ventilators manufactured and used in clinics. Therefore, the emphasis here is on the description of adult positive-pressure ventilators. Also, the concepts presented will be illustrated using a microprocessor-based design example, as almost all modern ventilators use microprocessor instrumentation.

26.4 Ventilation Modes

Since the advent of respirators, clinicians have devised a variety of strategies to ventilate the lungs based on patient conditions. For instance, some patients need the respirator to completely take over the task of ventilating their lungs. In this case, the ventilator operates in **mandatory mode** and delivers mandatory breaths. On the other hand, some patients are able to initiate a breath and breathe on their own, but may need oxygen-enriched air flow or slightly elevated airway pressure. When a ventilator assists a patient who is capable of demanding a breath, the ventilator delivers spontaneous breaths and operates in **spontaneous mode**. In many cases, it is first necessary to treat the patient with mandatory ventilation and as the patient's condition improves, spontaneous ventilation is introduced; it is used primarily to wean the patient from mandatory breathing.

Mandatory Ventilation

Designers of adult ventilators have employed two rather distinct approaches for delivering mandatory breaths: **volume-controlled ventilation** and **pressure-controlled ventilation**. Volume-controlled ventilation, which presently is more popular, refers to delivering a specified tidal volume to the patient during the inspiratory phase. Pressure-controlled ventilation, however, refers to raising the airway pressure to a level, set by the therapist, during the inspiratory phase of each breath. Regardless of the type, a ventilator operating in mandatory mode must control all aspects of breathing such as tidal volume, respiration rate, inspiratory flow pattern, and oxygen concentration of the breath. This is often labeled as **controlled mandatory ventilation (CMV)**.

Figure 26.3 shows the flow and pressure waveforms for a volume-controlled ventilation (CMV). In this illustration, the inspiratory flow waveform is chosen to be a half sinewave. In Fig. 26.3a, t_i is the inspiration duration, t_e is the exhalation period, and Q_i is the amplitude of inspiratory flow. The ventilator delivers a tidal volume equal to the area under the flow waveform in Fig. 26.3a at regular intervals $(t_i + t_e)$ set by the therapist. The resulting pressure waveform is shown in Fig. 26.3b. It is noted that during volume-controlled ventilation, the ventilator delivers the same volume irrespective of the patient's respiratory mechanics. However, the resulting pressure waveform such as the one shown in Fig. 26.3b, will be different among patients. Of course, for safety purposes, the ventilator limits the maximum applied airway pressure according to the therapist's setting.

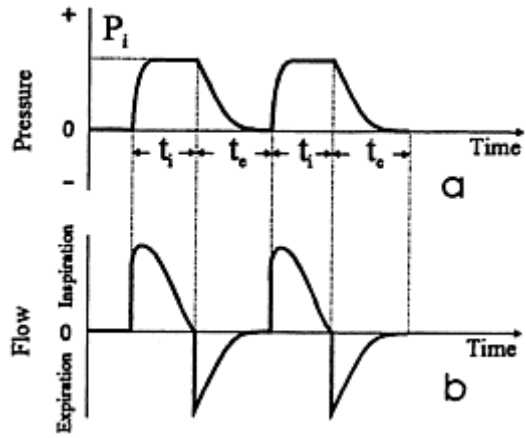


FIGURE 26.3 (a) Inspiratory flow for a controlled mandatory volume-controlled ventilation breath, (b) airway pressure resulting from the breath delivery with a non-zero PEEP.

As can be seen in Fig. 26.3b, the airway pressure at the end of exhalation may not end at atmospheric pressure (zero gauge). The **positive end expiratory pressure (PEEP)** is sometimes used to keep the alveoli from collapsing during expiration [Norwood, 1990]. In other cases, the expiration pressure is allowed to return to the atmospheric level.

Figure 26.4 shows a plot of the pressure and flow during a mandatory pressure-controlled ventilation. In this case, the respirator raises and maintains the airway pressure at the desired level independent of patient airway compliance and resistance. The level of pressure during inspiration, P_i, is set by the therapist. While the ventilator maintains the same pressure trajectory for patients with different respiratory resistance and compliance, the resulting flow trajectory, shown in Fig. 26.4b, will depend on the respiratory mechanics of each patient.

In the following, the presentation will focus on volume ventilators, as they are more common. Further, in a microprocessor-based ventilator, the mechanism for delivering mandatory volume and pressure controlled ventilation have many similar main components. The primary difference lies in the control algorithms governing the delivery of breaths to the patient.

Spontaneous Ventilation

An important phase in providing respiratory therapy to a recovering pulmonary patient is weaning the patient from the respirator. As the patient recovers and gains the ability to breathe independently, the ventilator must allow the patient to initiate a breath and control the breath rate, flow rate, and the tidal volume. Ideally, when a respirator is functioning in the spontaneous mode, it should let the patient take breaths with the same

ease as breathing from the atmosphere. This, however, is difficult to achieve because the respirator does not have an infinite gas supply or an instantaneous response. In practice, the patient generally has to exert more effort to breathe spontaneously on a respirator than from the atmosphere. However, patient effort is reduced as the ventilator response speed increases [McPherson, 1990]. Spontaneous ventilation is often used in conjunction with mandatory ventilation since the patient may still need breaths that are delivered entirely by the ventilator. Alternatively, when a patient can breathe completely on his own but needs oxygen-enriched breath or elevated airway pressure, spontaneous ventilation alone may be used.

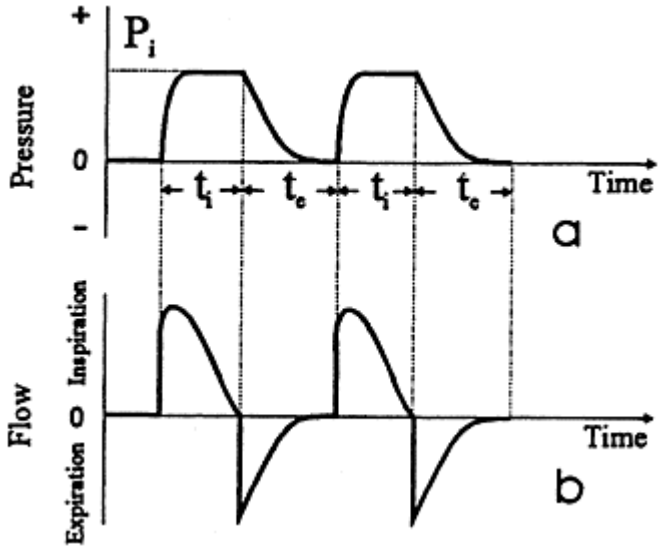


FIGURE 26.4 (a) Inspiratory pressure pattern for a controlled mandatory pressure-controlled ventilation breath, (b) airway flow pattern resulting from the breath delivery. Note that PEEP is zero.

As in the case of mandatory ventilation, several modes of spontaneous ventilation have been devised by therapists. Two of the most important and popular spontaneous breath delivery modes are described below.

Continuous Positive Airway Pressure (CPAP) in Spontaneous Mode

In this mode, the ventilator maintains a positive pressure at the airway as the patient attempts to inspire. Figure 26.5 illustrates a typical airway pressure waveform during CPAP breath delivery. The therapist sets the sensitivity level lower than PEEP. When the patient attempts to breathe, the pressure drops below the sensitivity level and the

ventilator responds by supplying breathable gases to raise the pressure back to the PEEP level. Typically, the PEEP and sensitivity levels are selected such that the patient will be impelled to exert effort to breathe independently. As in the case of the mandatory mode, when the patient exhales, the ventilator shuts off the flow of gas and opens the exhalation valve to allow the exhaled gases to flow into the atmosphere.

Pressure Support in Spontaneous Mode

This mode is similar to the CPAP mode with the exception that during the inspiration the ventilator attempts to maintain the patient airway pressure at a level above PEEP. In fact, CPAP may be considered a special case of pressure support ventilation in which the support level is fixed at the atmospheric level.

Figure 26.6 shows a typical airway pressure waveform during the delivery of a **pressure support** breath. In this mode, when the patient's airway pressure drops below the therapist-set sensitivity line, the

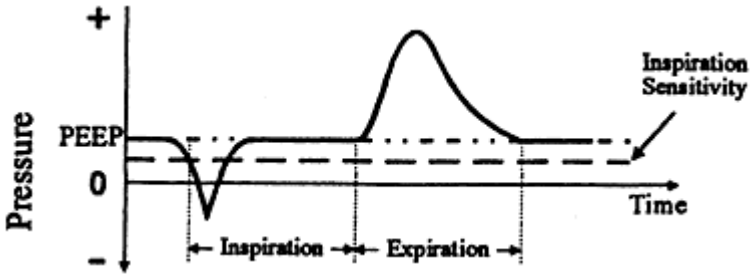


FIGURE 26.5 Airway pressure during a CPAP spontaneous breath delivery.

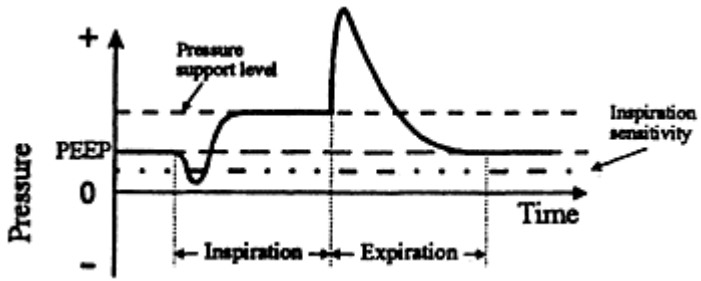


FIGURE 26.6 Airway pressure during a pressure support spontaneous breath delivery.

ventilator inspiratory breath delivery system raises the airway pressure to the pressure support level (>PEEP), selected by the therapist. The ventilator stops the flow of

breathable gases when the patient starts to exhale and controls the exhalation valve to achieve the set PEEP level.

26.5 Breath Delivery Control

Figure 26.7 shows a simplified block diagram for delivering mandatory or spontaneous ventilation. Compressed air and oxygen are normally stored in high-pressure tanks (1400 kPa) that are attached to the inlets of the ventilator. In some ventilators, an air compressor is used in place of a compressed air tank. Manufacturers of mechanical respirators have designed a variety of blending and metering devices [McPherson, 1990]. The primary mission of the device is to enrich the inspiratory air flow with the proper level of oxygen and to deliver a tidal volume according to the therapist's specifications. With the introduction of microprocessors for control of metering devices, electromechanical valves have gained popularity [Puritan-Bennett, 1987]. In Fig. 26.7, the air and oxygen valves are placed in closed feedback loops with the air and oxygen flow sensors. The microprocessor controls each of the valves to deliver the desired inspiratory air and oxygen flows for mandatory and spontaneous ventilation. During inhalation, the exhalation valve is closed to direct all the delivered flows to the lungs. When exhalation starts, the microprocessor actuates the exhalation valve to achieve the desired PEEP level. The airway pressure sensor, shown on the right side of Fig. 26.7, generates the feedback signal necessary for maintaining the desired PEEP (in both mandatory and spontaneous modes) and airway pressure support level during spontaneous breath delivery.

Mandatory Volume-Controlled Inspiratory Flow Delivery

In a microprocessor-controlled ventilator (Fig. 26.7), the electronically actuated valves open from a closed position to allow the flow of blended gases to the patient. The control of flow through each valve depends on the therapist's specification for the mandatory breath. That is, the clinician must specify the following parameters for delivery of CMV breaths: (1) respiration rate; (2) flow waveform; (3) tidal volume; (4) oxygen concentration (of the delivered breath); (5) peak flow; and (6) PEEP, as shown in the lower left side of Fig. 26.7. It is noted that the PEEP selected by the therapist in the mandatory mode is only used for control of exhalation flow; that will be described in the following section. The microprocessor utilizes the first five of the above parameters to compute the total desired inspiratory flow trajectory. To illustrate this point, consider the delivery of a tidal volume using a half sinewave as shown in Fig. 26.3. If the therapist selects a tidal volume of V_t (L), a respiration rate of n breaths per minute (bpm), the amplitude of the respirator flow, Q_i (L/s), then the total desired inspiratory flow, $Q_d(t)$, for a single breath, can be computed from the following equation:

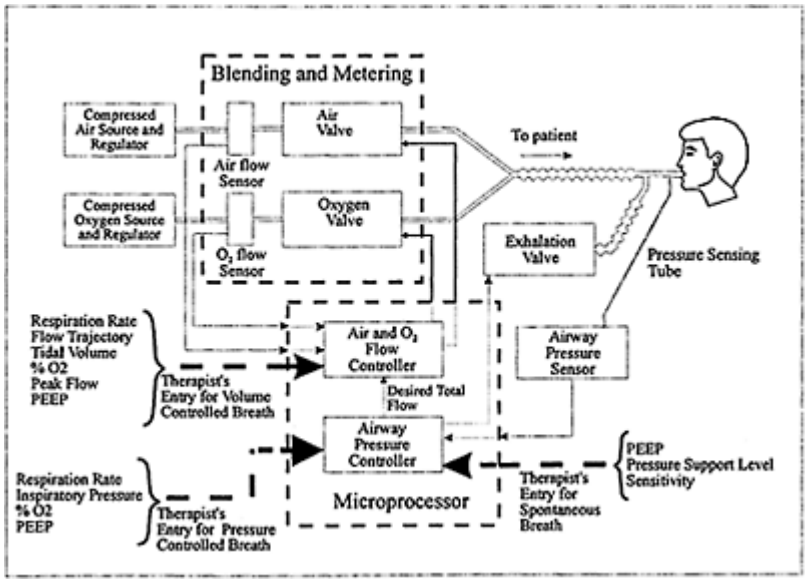


FIGURE 26.7 A simplified block diagram of a control structure for mandatory and spontaneous breath delivery.

$$Qd(t) = \begin{cases} Q_i \sin\dfrac{\pi t}{t_i} & o \le t < t_i \\ 0 & t_i < t \le t_e \end{cases}, \tag{26.1}$$

where t_i signifies the duration of inspiration and is computed from the following relationship:

$$t_i = \frac{V_t}{2Q_i}. \tag{26.2}$$

The duration of expiration in seconds is obtained from

$$t_e = \frac{60}{n} - t_i. \tag{26.3}$$

The ratio of inspiratory to expiratory periods of a mandatory breath is often used for adjusting the respiration rate. This ratio is represented by **I:E (ratio)** and is computed as follows. First, the inspiratory and expiratory periods are normalized with respect to t_i. Hence, the normalized inspiratory period becomes unity and the normalized expiratory period is given by $R=t_e/t_i$. Then, the I:E ratio is simply expressed as 1: R.

To obtain the desired oxygen concentration in the delivered breath, the microprocessor computes the discrete form of $Q_d(t)$ as $Q_d(k)$ where k signifies the kth sample interval. Then, the total desired flow, $Q_d(k)$, is partitioned using the following relationships:

$$Q_{da}(k) = \frac{(1-m)Q_d(k)}{(1-c)} \tag{26.4}$$

and

$$Q_{dx}(k) = \frac{(m-c)Q_d(k)}{(1-c)}, \tag{26.5}$$

where k signifies the sample interval, $Q_{da}(k)$ is the desired air flow (the subscript da stands for **d**esired air), $Q_{dx}(k)$ is the desired oxygen flow (the subscript dx stands for **d**esired oxygen), m is the desired oxygen concentration, and c is the oxygen concentration of the ventilator air supply.

A number of control design strategies may be appropriate for the control of the air and oxygen flow delivery valves. A simple controller is the proportional plus integral controller that can be readily implemented in a microprocessor. For example, the controller for the air valve has the following form:

$$I(k) = K_p E(k) + K_i A(k), \tag{26.6}$$

where $E(k)$ and $A(k)$ are given by

$$E(k) = Q_{da}(k) - Q_{sa}(k) \tag{26.7}$$

$$A(k) = A(k-1) + E(k), \tag{26.8}$$

where $I\ k)$ is the input (voltage or current) to the air valve at the kth sampling interval, $E(k)$ is the error in the delivered flow, $Q_{da}(k)$ is the desired air flow, $Q_{sa}(k)$ is the sensed or actual air flow (the subscript sa stands for sensed air flow), $A(k)$ is the integral (rectangular integration) part of the controller, and K_p and K_i are the controller proportionality constants. It is noted that the above equations are applicable to the control of either the air or oxygen valve. For control of the oxygen flow valve, $Q_{dx}(k)$ replaces $Q_{da}(k)$ and $Q_{sx}(k)$ replaces $Q_{sa}(k)$, where $Q_{sx}(k)$ represents the sensed oxygen flow (the subscript sx stands for sensed oxygen flow).

The control structure shown in Fig. 26.7 provides the flexibility of quickly adjusting the percentage of oxygen in the enriched breath gases. That is, the controller can regulate both the total flow and the percent oxygen delivered to the patient. Since the internal volume of the flow control valve is usually small (<50 ml), the desired change in the oxygen concentration of the delivered flow can be achieved within one inspiratory period. In actual clinical applications, rapid change of percent oxygen from one breath to another is often desirable, as it reduces the waiting time for the delivery of the desired

oxygen concentration. A design similar to the one shown in Fig. 26.7 has been successfully implemented in a microprocessor-based ventilator [Behbehani, 1984] and is deployed in hospitals around the world.

Pressure-Controlled Inspiratory Flow Delivery

The therapist entry for pressure-controlled ventilation is shown in Fig. 26.7 (lower left-hand side). In contrast to the volume-controlled ventilation where $Q_d(t)$ was computed directly from the operator's entry [Eq. (26.1) through Eq. (26.23)], the total desired flow is generated by a closed-loop controller labeled as Airway Pressure Controller in Fig. 26.7. This controller uses the therapist-selected inspiratory pressure, respiration rate, and the I:E ratio to compute the desired inspiratory pressure trajectory. The trajectory serves as the controller reference input. The controller then computes the flow necessary to make the actual airway pressure track the reference input. Assuming a proportional-plus-integral controller, the governing equations are

$$Q_d(k) = C_p E_p(k) + C_i A_p(k), \tag{26.9}$$

where Q_d is the computed desired flow, C_p and C_i are the proportionality constants, k represents the sample interval, and $E_p(k)$ and $A_p(k)$ are computed using the following equations:

$$E_p(k) = P_d(k) - P_s(k) \tag{26.10}$$

$$A_p(k) = A_p(k-1) + E_p(k), \tag{26.11}$$

where $E_p(k)$ is the difference between the desired pressure trajectory, $P_d(k)$, and the sensed airway pressure, $P_s(k)$, the parameter $A_p(k)$ represents the integral portion of the controller. Using Q_d from Eq. (26.9), thecontrol of air and O_2 valves is accomplished in the same manner as in the case of volume-controlledventilation described earlier [Eq. (26.4) through Eq. (26.8)].

Expiratory Pressure Control in Mandatory Mode

It is often desirable to keep the patient's lungs inflated at the end of expiration at a pressure greater than atmospheric level [Norwood, 1990]. That is, rather than allowing the lungs to deflate during the exhalation, the controller closes the exhalation valve when the airway pressure reaches the PEEP level. When expiration starts, the ventilator terminates flow to the lungs; hence, the regulation of the airway pressure is achieved by controlling the flow of patient exhaled gases through the exhalation valve.

In a microprocessor-based ventilator, an electronically actuated valve can be employed that has adequate dynamic response (Φ20 ms rise time) to regulate PEEP. For this purpose, the pressure in the patient breath delivery circuit is measured using a pressure

transducer (Fig. 26.7). The microprocessor will initially open the exhalation valve completely to minimize resistance to expiratory flow. At the same time, it will sample the pressure transducer's output and start to close the exhalation valve as the pressure begins to approach the desired PEEP level. Since the patient's exhaled flow is the only source of pressure, if the airway pressure drops below PEEP, it cannot be brought back up until the next inspiratory period. Hence, an overrun (i.e., a drop to below PEEP) in the closed-loop control of PEEP cannot be tolerated.

Spontaneous Breath Delivery Control

The small diameter (Φ5 mm) pressure sensing tube, shown on the right side of Fig. 26.7, pneumatically transmits the pneumatic pressure signal from the patient airway to a pressure transducer placed in the ventilator. The output of the pressure transducer is amplified, filtered, and then sampled by the microprocessor. The controller receives the therapist's inputs regarding the spontaneous breath characteristics such as the PEEP, sensitivity, and oxygen concentration, as shown on the lower right-hand side of Fig. 26.7. The desired airway pressure is computed from the therapist entries of PEEP, pressure support level, and sensitivity. The multiple-loop control structure shown in Fig. 26.7 is used to deliver a CPAP or a pressure support breath. The sensed proximal airway pressure is compared with the desired airway pressure. The airway pressure controller computes the total inspiratory flow level required to raise the airway pressure to the desired level. This flow level serves as the reference input or total desired flow for the flow control loop. Hence, in general, the desired total flow trajectory for the spontaneous breath delivery may be different for each inspiratory cycle. If the operator has specified oxygen concentration greater than 21.6% (the atmospheric air oxygen concentration of the ventilator air supply), the controller will partition the total required flow into the air and oxygen flow rates using Eqs. (26.4) and (26.5). The flow controller then uses the feedback signals from air and oxygen flow sensors and actuates the air and oxygen valves to deliver the desired flows.

For a microprocessor-based ventilator, the control algorithm for regulating the airway pressure can also be a proportional plus integral controller [Behbehani, 1984; Behbehani and Watanabe, 1986]. In this case, the governing equations are identical to Eqs. (26.9) through (26.11).

If a nonzero PEEP level is specified, the same control strategy as the one described for mandatory breath delivery can be used to achieve the desired PEEP.

26.6 Summary

Today's mechanical ventilators can be broadly classified into negative-pressure and positive-pressure ventilators. Negative-pressure ventilators do not offer the flexibility and convenience that positive-pressure ventilators provide; hence, they have not been very popular in clinical use. Positive-pressure ventilators have been quite successful in treating patients with pulmonary disorders. These ventilators operate in either mandatory or spontaneous mode. When delivering mandatory breaths, the ventilator controls all parameters of the breath such as tidal volume, inspiratory flow waveform, respiration

rate, and oxygen content of the breath. Mandatory breaths are normally delivered to the patients that are incapable of breathing on their own. In contrast, spontaneous breath delivery refers to the case where the ventilator responds to the patient's effort to breathe independently. Therefore, the patient can control the volume and the rate of the respiration. The therapist selects the oxygen content and the pressure at which the breath is delivered. Spontaneous breath delivery is typically used for patients who are on their way to full recovery, but are not completely ready to breathe from the atmosphere without mechanical assistance.

Defining Terms

Continuous positive airway pressure (CPAP): A spontaneous ventilation mode in which the ventilator maintains a constant positive pressure, near or below PEEP level, in the patient's airway while the patient breathes at will.
I: E ratio: The ratio of normalized inspiratory interval to normalized expiratory interval of a mandatory breath. Both intervals are normalized with respect to the inspiratory period. Hence, the normalized inspiratory period is always unity.
Mandatory mode: A mode of mechanically ventilating the lungs where the ventilator controls all breath delivery parameters such as tidal volume, respiration rate, flow waveform, etc.
Patient circuit: A set of tubes connecting the patient airway to the outlet of a respirator.
Positive end expiratory pressure (PEEP): A therapist-selected pressure level for the patient airway at the end of expiration in either mandatory or spontaneous breathing.
Pressure controlled ventilation: A mandatory mode of ventilation where during the inspiration phase of each breath, a constant pressure is applied to the patient's airway independent of the patient's airway resistance and/or compliance.respiratory mechanics.
Pressure support: A spontaneous breath delivery mode during which the ventilator applies a positive pressure greater than PEEP to the patient's airway during inspiration.
Pressure support level: Refers to the pressure level, above PEEP, that the ventilator maintains during the spontaneous inspiration.
Spontaneous mode: A ventilation mode in which the patient initiates and breathes from the ventilator-supplied gas at will.
Volume-controlled ventilation: A mandatory mode of ventilation where the volume of each breath is set by the therapist and the ventilator delivers that volume to the patient independent of the patient's airway resistance and/or compliance respiratory mechanics.

References

Puritan-Bennett 7200 Ventilator System Series, "Ventilator, Options and Accessories," Part. No. 22300A, Carlsbad, CA, September 1990.

Pilbeam, S.P., 1998. *Mechanical Ventilation: Physiological and Clinical Applications,* 3rd ed, Mosby, St. Louis, MO.

McPherson, S.P. and Spearman, C.B. 1990. *Respiratory Therapy Equipment,* 4th ed. C.V.Mosby Co., St. Louis, MO.

Norwood, S. 1990. Physiological principles of conventional mechanical ventilation. In *Clinical Application of Ventilatory Support,* Kirby, R.R., Banner, M.J., and Downs, J.B., Eds., Churchill Livingstone, New York, 145–172.

Behbehani, K. 1984. PLM-implementation of a multiple closed-loop control strategy for a microprocessor-controlled respirator. *Proc. ACC Conf.,* p. 574–576.

Behbehani, K. and Watanabe, N.T. 1986. A new application of digital computer simulation in the design of a microprocessor-based respirator. *Summer Simulation Conf.,* p. 415–420.

27
Parenteral Infusion Devices

Gregory I.Voss
IVAC Corporation

Robert D.Butterfield
IVAC Corporation

The circulatory system is the body's primary pathway for both the distribution of oxygen and other nutrients and the removal of carbon dioxide and other waste products. Since the entire blood supply in a healthy adult completely circulates within 60 seconds, substances introduced into the circulatory system are distributed rapidly. Thus intravenous (IV) and intraarterial access routes provide an effective pathway for the delivery of fluid, blood, and medicants to a patient's vital organs. Consequently, about 80% of hospitalized patients receive infusion therapy. Peripheral and central veins are used for the majority of infusions. Umbilical artery delivery (in neonates), enteral delivery of nutrients, and epidural delivery of anesthetics and analgesics comprise smaller patient populations. A variety of devices can be used to provide flow through an intravenous catheter. An intravenous delivery system typically consists of three major components: (1) fluid or drug reservoir, (2) catheter system for transferring the fluid or drug from the reservoir into the vasculature through a venipuncture, and (3) device for regulation and/or generating flow (see Fig. 27.1).

This chapter is separated into five sections. The first describes the clinical needs associated with intravenous drug delivery that determine device performance criteria. The second section reviews the principles of flow through a tube; the third section introduces the underlying electromechanical principles for flow regulation and/or generation and their ability to meet the clinical performance criteria. The fourth section reviews complications associated with intravenous therapy, and the fifth section concludes with a short list of articles providing more detailed information.

27.1 Performance Criteria for Intravenous Infusion Devices

The intravenous pathway provides an excellent route for continuous drug therapy. The ideal delivery system regulates drug concentration in the body to achieve and maintain a desired result. When the drug's effect cannot be monitored directly, it is frequently

0-8493-1813-0/03/$0.00+$ 1.50

assumed that a specific blood concentration or infusion rate will achieve the therapeutic objective. Although underinfusion may not provide sufficient therapy, overinfusion can produce even more serious toxic side effects.

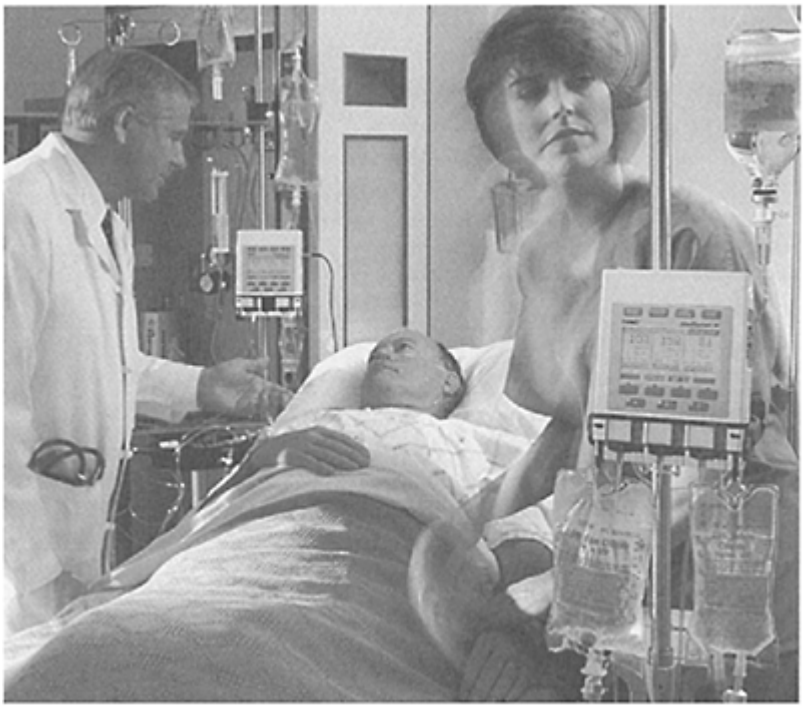


FIGURE 27.1 A typical IV infusion system.

The therapeutic range and risks associated with under- and overinfusion are highly drug and patient dependent. Intravenous delivery of fluids and electrolytes often does not require very accurate regulation. Low-risk patients can generally tolerate well infusion rate variability of ±30% for fluids. In some situations, however, specifically for fluid-restricted patients, prolonged under- or overinfusion of fluids can compromise the patient's cardiovascular and renal systems.

The infusion of many drugs, especially potent cardioactive agents, requires high accuracy. For example, postcoronary-artery-bypass-graft patients commonly receive sodium nitroprusside to lower arterial blood pressure. Hypertension, associated with underinfusion, subjects the graft sutures to higher stress with an increased risk for internal bleeding. Hypotension associated with overinfusion can compromise the cardiovascular state of the patient. Nitroprusside's potency, short onset delay, and short half-life (30 to 180 s) provide for very tight control, enabling the clinician to quickly respond to the many events that alter the patient's arterial pressure. The fast response of drugs such as nitroprusside creates a need for short-term flow uniformity as well as long-term accuracy.

The British Department of Health employs *trumpet curves* in their Health Equipment Information reports to compare flow uniformity of infusion pumps. For a prescribed flow rate, the trumpet curve is the plot of the maximum and minimum measured percentage flow rate error as a function of the accumulation interval (Fig. 27.2). Flow is measured gravimetrically in 30-s blocks for 1 hour. These blocks are summed to produce 120-s, 300-s, and other longer total accumulation intervals. Though the 120-s window may not detect flow variations important in delivery of the fastest acting agents, the trumpet curve provides a helpful means for performance comparison among infusion devices. Additional statistical information such as standard deviations may be derived from the basic trumpet flow measurements.

The short half-life of certain pharmacologic agents and the clotting reaction time of blood during periods of stagnant flow require that fluid flow be maintained without significant interruption. Specifically, concern has been expressed in the literature that the infusion of sodium nitroprusside and other short half-life drugs occur without interruption exceeding 20 *s*. Thus, minimization of false alarms and rapid detection of occlusions are important aspects of maintaining a constant vascular concentration.

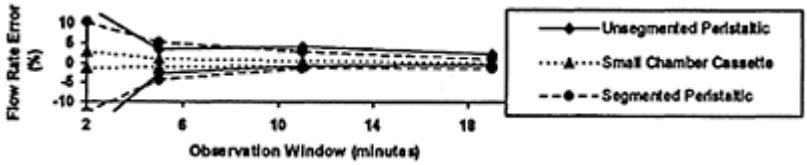


FIGURE 27.2 Trumpet curve for several representative large volume infusion pumps operated at 5 mL/hr. Note that peristaltic pumps were designed for low-risk patients.

Accidental occlusions of the IV line due to improper positioning of stopcocks or clamps, kinked tubing, and clotted catheters are common.

Occlusions between pump and patient present a secondary complication in maintaining serum drug concentration. Until detected, the pump will infuse, storing fluid in the delivery set. When the occlusion is eliminated, the stored volume is delivered to the patient in a bolus. With concentrated pharmaceutic agents, this bolus can produce a large perturbation in the patient's status.

Occlusions of the pump intake also interrupt delivery. If detection is delayed, inadequate flow can result. During an intake occlusion, in some pump designs removal of the delivery set can produce abrupt aspiration of blood. This event may precipitate clotting and cause injury to the infusion site.

The common practice of delivering multiple drugs through a single venous access port produces an additional challenge to maintaining uniform drug concentration. Although some mixing will occur in the venous access catheter, fluid in the catheter more closely resembles a first-in/first-out digital queue: During delivery, drugs from the various infusion devices mix at the catheter input, an equivalent fluid volume discharges from the

outlet. Rate changes and flow nonuniformity cause the mass flow of drugs at the outlet to differ from those at the input. Consider a venous access catheter with a volume of 2 mL and a total flow of 10 mL/hr. Due to the digital queue phenomenon, an incremental change in the intake flow rate of an individual drug will not appear at the output for 12 min. In addition, changing flow rates for one drug will cause short-term marked swings in the delivery rate of drugs using the same access catheter. When the delay becomes significantly larger than the time constant for a drug that is titrated to a measurable patient response, titration becomes extremely difficult leading to large oscillations.

As discussed, the performance requirements for drug delivery vary with multiple factors: drug, fluid restriction, and patient risk. Thus the delivery of potent agents to fluid-restricted patients at risk require the highest performance standards defined by flow rate accuracy, flow rate uniformity, and ability to minimize risk of IV-site complications. These performance requirements need to be appropriately balanced with the device cost and the impact on clinician productivity.

27.2 Flow Through an IV Delivery System

The physical properties associated with the flow of fluids through cylindrical tubes provide the foundation for understanding flow through a catheter into the vasculature. Hagen-Poiseuille's equation for laminar flow of a Newtonian fluid through a rigid tube states:

$$Q = \pi \cdot r^4 \cdot \frac{\left(P_1 - P_2\right)}{8 \cdot \eta \cdot L},$$

where Q is the flow; P_1 and P_2 are the pressures at the inlet and outlet of the tube, respectively; L and r are the length and internal radius of the tube, respectively; and η is fluid viscosity. Although many drug delivery systems do not strictly meet the flow conditions for precise application of the laminar flow equation, it does provide insight into the relationship between flow and pressure in a catheter. The fluid analog of Ohms law describes the resistance to flow under constant flow conditions:

TABLE 27.1 Resistance Measurements for Catheter Components Used for Infusion

Component	Length, cm	Flow Resistance, Fluid Ohm, mmHg/(L/hr)
Standard administration set	91–213	4.3–5.3
Extension tube for CVP monitoring	15	15.5
19-gauge epidural catheter	91	290.4–497.1
18-gauge needle	6–9	14.1–17.9
23-gauge needle	2.5–9	165.2–344.0

25-gauge needle	1.5–4.0	525.1–1412.0
Vicra Quick-Cath Catheter 18-gauge	5	12.9
Extension set with 0.22 micron air-eliminating filter		623.0
0.2 micron filter		555.0

Note: Mean values are presented over a range of infusions (100, 200, and 300 mL/hr) and sample size (n=10).

$$R = \frac{P_1 - P_2}{Q}.$$

Thus, resistance to flow through a tube correlates directly with catheter length and fluid viscosity and inversely with the fourth power of catheter diameter. For steady flow, the delivery system can be modeled as a series of resistors representing each component, including administration set, access catheter, and circulatory system. When dynamic aspects of the delivery system are considered, a more detailed model including catheter and venous compliance, fluid inertia, and turbulent flow is required. Flow resistance may be defined with units of mmHg/(L/hr), so that 1 fluid ohm=4.8×10^{-11} Pa s/m^3. Studies determining flow resistance for several catheter components with distilled water for flow rates of 100,200, and 300 mL/ hr appear in Table 27.1.

27.3 Intravenous Infusion Devices

From Hagen-Poiselluie's equation, two general approaches to intravenous infusion become apparent. First, a hydrostatic pressure gradient can be used with adjustment of delivery system resistance controlling flow rate. Complications such as partial obstructions result in reduced flow that may be detected by an automatic flow monitor. Second, a constant displacement flow source can be used. Now complications may be detected by monitoring elevated fluid pressure and/or flow resistance. At the risk of overgeneralization, the relative strengths of each approach will be presented.

Gravity Flow/Resistance Regulation

The simplest means for providing regulated flow employs gravity as the driving force with a roller clamp as controlled resistance. Placement of the fluid reservoir 60 to 100 cm above the patient's right atrium provides a hydrostatic pressure gradient P_h equal to 1.34 mmHg per cm of elevation. The modest physiologic mean pressure in the veins, P_v, minimally reduces the net hydrostatic pressure gradient. The equation for flow becomes

$$Q = \frac{P_h - P_v}{R_{mfr} + R_n},$$

where R_{mfr} and R_n are the resistance to flow through the mechanical flow regulator and the remainder of the delivery system, respectively. Replacing the variables with representative values for an infusion of 5% saline solution into a healthy adult at 100 mL/hr yields

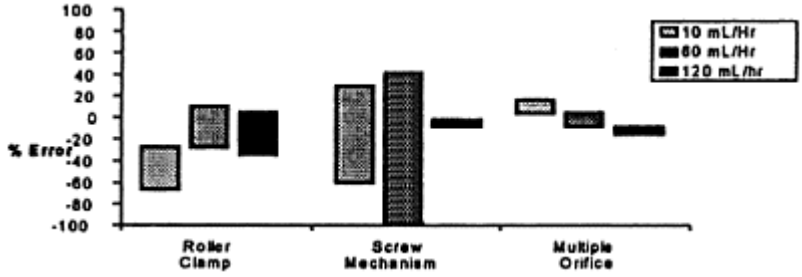


FIGURE 27.3 Drift in flow rate (mean±standard deviation) over a four-hour period for three mechanical flow regulators at initial flow rates of 10, 60, and 120 mL/hr with distilled water at constant hydrostatic pressure gradient.

$$100\,mL/hr = \frac{(68-8)\,mmHg}{(550+50)\,\frac{mmHg}{(L/hr)}}$$

Gravity flow cannot be used for arterial infusions since the higher vascular pressure exceeds available hydrostatic pressure.

Flow stability in a gravity infusion system is subject to variations in hydrostatic and venous pressure as well as catheter resistance. However, the most important factor is the change in flow regulator resistance caused by viscoelastic creep of the tubing wall (see Fig. 27.3). Caution must be used in assuming that a preset flow regulator setting will accurately provide a predetermined rate. The clinician typically estimates flow rate by counting the frequency of drops falling through an in-line drip-forming chamber, adjusting the clamp to obtain the desired drop rate. The cross-sectional area of the drip chamber orifice is the major determinant of drop volume. Various manufacturers provide minidrip sets designed for pediatric (e.g., 60 drops/mL) and regular sets designed for adult (10 to 20 drops/mL) patients. Tolerances on the drip chamber can cause a 3% error in minidrip sets and a 17% error in regular sets at 125 mL/hr flow rate with 5% dextrose in water. Mean drop size for rapid rates increased by as much as 25% over the size of drops that form slowly. In addition, variation in the specific gravity and surface tension of fluids can provide an additional large source of drop size variability.

Some mechanical flow regulating devices incorporate the principle of a Starling resistor. In a Starling device, resistance is proportional to hydrostatic pressure gradient. Thus, the device provides a negative feedback mechanism to reduce flow variation as the available pressure gradient changes with time.

Mechanical flow regulators comprise the largest segment of intravenous infusion systems, providing the simplest means of operation. Patient transport is simple, since these devices require no electric power. Mechanical flow regulators are most useful where the patient is not fluid restricted and the acceptable therapeutic rate range of the drug is relatively wide with minimal risk of serious adverse sequelae. The most common use for these systems is the administration of fluids and electrolytes.

Volumetric Infusion Pumps

Active pumping infusion devices combine electronics with a mechanism to generate flow. These devices have higher performance standards than simple gravity-flow regulators. The Association for the Advancement of Medical Instrumentation (AAMI) recommends that long-term rate accuracy for infusion pumps remain within ±10% of the set rate for general infusion and, for the more demanding applications, that long-term flow remain within ±5%. Such requirements typically extend to those agents with narrow therapeutic indices and/or low flow rates, such as the neonatal population or other fluid-restricted patients. The British Department of Health has established three main categories for hospital-based infusion devices: neonatal infusions, high-risk infusions, and low-risk infusions. Infusion control for neonates requires the highest performance standards, because their size severely restricts fluid volume. A fourth category, ambulatory infusion, pertains to pumps worn by patients.

Controllers

These devices automate the process of adjusting the mechanical flow regulator. The most common controllers utilize sensors to count the number of drops passing through the drip chamber to provide flow feedback for automatic rate adjustment. Flow rate accuracy remains limited by the rate and viscosity dependence of drop size. Delivery set motion associated with ambulation and improper angulation of the drip chamber can also hinder accurate rate detection.

An alternative to the drop counter is a volumetric metering chamber. A McGaw Corporation controller delivery set uses a rigid chamber divided by a flexible membrane. Instrument-controlled valves allow fluid to fill one chamber from the fluid reservoir, displacing the membrane driving the fluid from the second chamber toward the patient. When inlet and outlet valves reverse state, the second chamber is filled while the first chamber delivers to the patient. The frequency of state change determines the average flow rate. Volumetric accuracy demands primarily on the dimensional tolerances of the chamber. Although volumetric controllers may provide greater accuracy than drop-counting controllers, their disposables are inherently more complex, and maximum flow is still limited by head height and system resistance.

Beyond improvements in flow rate accuracy, controllers should provide an added level of patient safety by quickly detecting IV-site complications. The IVAC Corporation has

developed a series of controllers employing pulsed modulated flow providing for monitoring of flow resistance as well as improved accuracy.

The maximum flow rate achieved by gravimetric-based infusion systems can become limited by R_n and by concurrent infusion from other sources through the same catheter. In drop-counting devices, flow rate uniformity suffers at low flow rates from the discrete nature of the drop detector.

In contrast with infusion controllers, pumps generate flow by mechanized displacement of the contents of a volumetric chamber. Typical designs provide high flow-rate accuracy and uniformity for a wide rate range (0.1 to 1000.0 mL/hr) of infusion rates. Rate error correlates directly with effective chamber volume, which, in turn, depends on both instrument and disposable repeatability, precision, and stability under varying load. Stepper or servo-controlled dc motors are typically used to provide the driving force for the fluid. At low flow rates, dc motors usually operate in a discrete stepping mode. On average, each step propels a small quanta of fluid toward the patient. Flow-rate uniformity therefore is a function of both the average volume per quanta and the variation in volume. Mechanism factors influencing rate uniformity include: stepping resolution, gearing and activator geometries, volumetric chamber coupling geometry, and chamber elasticity. When the quanta volume is not inherently uniform over the mechanism's cycle, software control has been used to compensate for the variation.

Syringe Pumps

These pumps employ a syringe as both reservoir and volumetric pumping chamber. A precision leadscrew is used to produce constant linear advancement of the syringe plunger. Except for those ambulatory systems that utilize specific microsyringes, pumps generally accept syringes ranging in size from 5 to 100 mL. Flow rate accuracy and uniformity are determined by both mechanism displacement characteristics and tolerance on the internal syringe diameter. Since syringe mechanisms can generate a specified linear travel with less than 1% error, the manufacturing tolerance on the internal cross-sectional area of the syringe largely determines flow rate accuracy. Although syringes can be manufactured to tighter tolerances, standard plastic syringes provide long-term accuracy of ±5%. Flow rate uniformity, however, can benefit from the ability to select syringe size (see Fig. 27.4). Since many syringes have similar stroke length, diameter variation provides control of volume. Also the linear advancement per step is typically fixed. Therefore selection of a lower-volume syringe provides smaller-volume quanta. This allows tradeoffs among drug concentration,

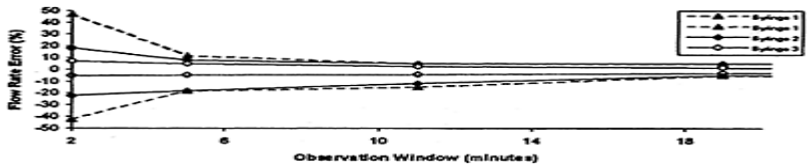


FIGURE 27.4 Effect of syringe type on trumpet curve of a syringe pump at 1 mL/hr.

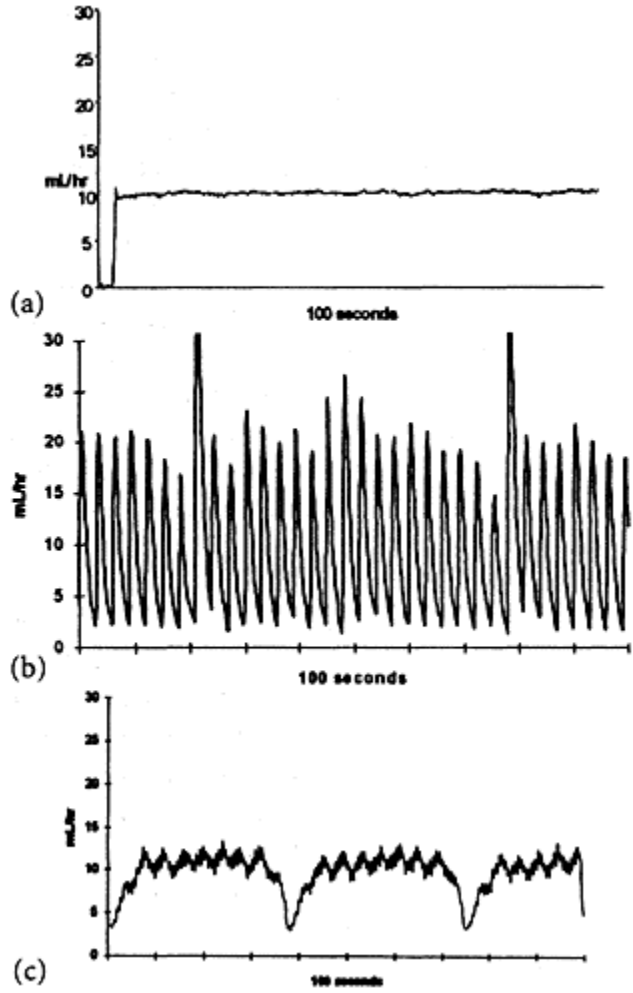


FIGURE 27.5 Continuous flow pattern for a representative, *(a)* syringe, *(b)* cassette, and *(c)* linear peristaltic pump at 10 mL/hr.

flow rate, and duration of flow per syringe. Slack in the gear train and drive shaft coupling as well as plunger slip cause rate inaccuracies during the initial stages of delivery (see Fig. 27.5*a*).

Since the syringe volumes are typically much smaller than reservoirs used with other infusion devices, syringe pumps generally deliver drugs in either fluid-restricted environments or for short duration. With high-quality syringes, flow-rate uniformity in syringe pumps is generally superior to that accomplished by other infusion pumps. With

the drug reservoir enclosed within the device, syringe pumps manage patient transport well, including the operating room environment.

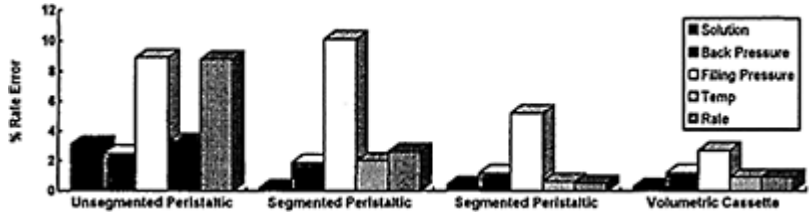


FIGURE 27.6 Impact of 5 variables on flow rate accuracy in 4 different infusion pumps. Variables tested included solution: distilled water and 25% dextrose in water; back pressure: −100 mmHg and 300 mmHg; pumping segment filling pressure: −30 inches of water and +30 inches of water; temperature: 10°C and 40°C; and infusion rate: 5 mL/ hr and 500 mL/hr. Note: First and second peristaltic mechanism qualified for low-risk patients, while the third peristaltic device qualified for high-risk patients.

Cassette pumps conceptually mimic the piston type action of the syringe pump but provide an automated means of repeatedly emptying and refilling the cassette. The process of refilling the cassette in single piston devices requires an interruption in flow (see Fig. 27.5*b*). The length of interruption relative to the drug's half-life determines the impact of the refill period on hemodynamic stability. To eliminate the interruption caused by refill, dual piston devices alternate refill and delivery states, providing nearly continuous output. Others implement cassettes with very small volumes that can refill in less than a second (see Fig. 27.2). Tight control of the internal cross-sectional area of the pumping chamber provides exceptional flow rate accuracy. Manufacturers have recently developed remarkably small cassette pumps that can still generate the full spectrum of infusion rate (0.1 to 999.0 mL/hr). These systems combine pumping chamber, inlet and outlet valving, pressure sensing, and air detection into a single complex component.

Peristaltic pumps operate on a short segment of the IV tubing. Peristaltic pumps can be separated into two subtypes. Rotary peristaltic mechanisms operate by compressing the pumping segment against the rotor housing with rollers mounted on the housing. With rotation, the rollers push fluid from the container through the tubing toward the patient. At least one of the rollers completely occludes the tubing against the housing at all times precluding free flow from the reservoir to the patient. During a portion of the revolution,

two rollers trap fluid in the intervening pumping segment. The captured volume between the rollers determines volumetric accuracy. Linear peristaltic pumps hold the pumping segment in a channel pressed against a rigid backing plate. An array of cam-driven actuators sequentially occlude the segment starting with the section nearest the reservoir forcing fluid toward the patient with a sinusoidal wave action. In a typical design using uniform motor step intervals, a characteristic flowwave resembling a positively biased sinewave is produced (see Fig. 27.5*c*).

Infusion pumps provide significant advantages over both manual flow regulators and controllers in several categories. Infusion pumps can provide accurate delivery over a wide range of infusion rates (0.1 to 999.0 mL/hr). Neither elevated system resistance nor distal line pressure limit the maximum infusion rate. Infusion pumps can support a wider range of applications including arterial infusions, spinal and epidural infusions, and infusions into pulmonary artery or central venous catheters. Flow rate accuracy of infusion pumps is highly dependent on the segment employed as the pumping chamber (see Fig. 27.2). Incorporating special syringes or pumping segments can significantly improve flow rate accuracy (see Fig. 27.6). Both manufacturing tolerances and segment material composition significantly dictate flow rate accuracy. Time-and temperature-related properties of the pumping segment further impact long-term drift in flow rate.

27.4 Managing Occlusions of the Delivery System

One of the most common problems in managing an IV delivery system is the rapid detection of occlusion in the delivery system. With a complete occlusion, the resistance to flow approaches infinity. In this condition, gravimetric-based devices cease to generate flow. Mechanical flow regulators have no mechanism for adverse event detection and thus must rely on the clinician to identify an occlusion as part of routine patient care. Electronic controllers sense the absence of flow and alarm in response to their inability to sustain the desired flow rate.

The problem of rapidly detecting an occlusion in an infusion pump is more complex. Upstream occlusions that occur between the fluid reservoir and the pumping mechanism impact the system quite differently than downstream occlusions that occur between the pump and the patient. When an occlusion occurs downstream from an infusion pump, the pump continues to propel fluid into the section of tubing between the pump and the occlusion. The time rate of pressure rise in that section increases in direct proportion to flow rate and inversely with tubing compliance (compliance, C, is the volume increase in a closed tube per mmHg pressure applied). The most common approach to detecting downstream occlusion requires a pressure transducer immediately below the pumping mechanism. These devices generate an alarm when either the mean pressure or rate of change in pressure exceeds a threshold. For pressure-limited designs, the time to downstream alarm (TTA) may be estimated as

$$TTA = \frac{P_{alarm} \cdot C_{delivery-set}}{flow\ rate}.$$

Using a representative tubing compliance of 1 μL/mmHg, flow rate of 1 mL/hr, and a fixed alarm threshold set of 500 mmHg, the time to alarm becomes

$$TTA = \frac{500_{\text{mmHg}} \cdot 1000\,\text{ml}/\text{mmHg}}{1\,\text{mL}/\text{hr}} = 30\ \text{min},$$

where TTA is the time from occlusion to alarm detection. Pressure-based detection algorithms depend on accuracy and stability of the sensing system. Lowering the threshold on absolute or relative pressure for occlusion alarm reduces the TTA, but at the cost of increasing the likelihood of false alarms. Patient movement, patient-to-pump height variations, and other clinical circumstances can cause wide perturbations in line pressure. To optimize the balance between fast TTA and minimal false alarms, some infusion pumps allow the alarm threshold to be set by the clinician or be automatically shifted upward in response to alarms; other pumps attempt to optimize performance by varying pressure alarm thresholds with flow rate.

A second approach to detection of downstream occlusions uses motor torque as an indirect measure of the load seen by the pumping mechanism. Although this approach eliminates the need for a pressure sensor, it introduces additional sources for error including friction in the gear mechanism or pumping mechanism that requires additional safety margins to protect against false alarms. In syringe pumps, where the coefficient of static friction of the syringe bunge (rubber end of the syringe plunger) against the syringe wall can be substantial, occlusion detection can exceed 1 hr at low flow rates.

Direct, continuous measurement of downstream flow resistance may provide a monitoring modality that overcomes the disadvantages of pressure-based alarm systems, especially at low infusion rates. Such a monitoring system would have the added advantage of performance unaffected by flow rate, hydrostatic pressure variations, and motion artifacts.

Upstream occlusions can cause large negative pressures as the pumping mechanism generates a vacuum on the upstream tubing segment. The tube may collapse and the vacuum may pull air through the tubing walls or form cavitation bubbles. A pressure sensor situated above the mechanism or a pressure sensor below the mechanism synchronized with filling of the pumping chamber can detect the vacuum associated with an upstream occlusion. Optical or ultrasound transducers, situated below the mechanism, can detect air bubbles in the catheter, and air-eliminating filters can remove air, preventing large air emboli from being introduced into the patient.

Some of the most serious complications of IV therapy occur at the venipuncture site; these include extravasation, postinfusion phlebitis (and thrombophlebitis), IV-related infections, ecchymosis, and hematomas. Other problems that do not occur as frequently include speed shock and allergic reactions.

Extravasation (or infiltration) is the inadvertent perfusion of infusate into the interstitial tissue. Reported percentage of patients to whom extravasation has occurred ranges from 10% to over 25%. Tissue damage does not occur frequently, but the consequences can be severe, including skin necrosis requiring significant plastic and reconstructive surgery and amputation of limbs. The frequency of extravasation injury correlates with age, state of consciousness, and venous circulation of the patient as well as the type, location, and placement of the intravenous cannula. Drugs that have high

osmolality, vessicant properties, or the ability to induce ischemia correlate with frequency of extravasation injury. Neonatal and pediatric patients who possess limited communication skills, constantly move, and have small veins that are difficult to cannulate require superior vigilance to protect against extravasation.

Since interstitial tissue provides a greater resistance to fluid flow than the venous pathway, infusion devices with accurate and precise pressure monitoring systems have been used to detect small pressure increases due to extravasation. To successfully implement this technique requires diligence by the clinician, since patient movement, flow rate, catheter resistance, and venous pressure variations can obscure the small pressure variations resulting from the extravasation. Others have investigated the ability of a pumping mechanism to withdraw blood as indicative of problems in a patent line. The catheter tip, however, may be partially in and out of the vein such that infiltration occurs yet blood can be withdrawn from the patient. A vein might also collapse under negative pressure in a patient line without successful blood withdrawal. Techniques currently being investigated that monitor infusion impedance (resistance and compliance) show promise for assisting in the detection of extravasation.

When a catheter tip wedges into the internal lining of the vein wall, it is considered positional. With the fluid path restricted by the vein wall, increases in line resistance may indicate a positional catheter. With patient movement, for example wrist flexation, the catheter may move in and out of the positional state. Since a positional catheter is thought to be more prone toward extravasation than other catheters, early detection of a positional catheter and appropriate adjustment of catheter position may be helpful in reducing the frequency of extravasation.

Postinfusion phlebitis is acute inflammation of a vein used for IV infusion. The chief characteristic is a reddened area or red streak that follows the course of the vein with tenderness, warmth, and edema at the venipuncture site. The vein, which normally is very compliant, also hardens. Phlebitis positively correlates with infusion rate and with the infusion of vesicants.

Fluid overload and speed shock result from the accidental administration of a large fluid volume over a short interval. Speed shock associates more frequently with the delivery of potent medications, rather than fluids. These problems most commonly occur with manually regulated IV systems, which do not provide the safety features of instrumented lines. Many IV sets designed for instrumented operation will free-flow when the set is removed from the instrument without manual clamping. To protect against this possibility, some sets are automatically placed in the occluded state on disengagement. Although an apparent advantage, reliance on such automatic devices may create a false sense of security and lead to manual errors with sets not incorporating these features.

27.5 Summary

Intravenous infusion has become the mode of choice for delivery of a large class of fluids and drugs both in hospital and alternative care settings. Modern infusion devices provide the clinician with a wide array of choices for performing intravenous therapy. Selection of the appropriate device for a specified application requires understanding of drug

pharmacology and pharmacokinetics, fluid mechanics, and device design and performance characteristics. Continuing improvements in performance, safety, and cost of these systems will allow even broader utilization of intravenous delivery in a variety of settings.

References

Association for the Advancement of Medical Instrumentation. 1992. Standard for Infusion Devices. Arlington, Virg.

Bohony J. 1993. Nine common intravenous complications and what to do about them. *Am J Nursing* 10:45.

British Department of Health. 1990. *Evaluation of Infusion Pumps and Controllers.* HEI Report #198.

Glass PSA, Jacobs JR, Reves JG. 1991. Technology for continuous infusions in anesthesia. Continuous Infusions in Anesthesia. *International Anesthesiology Clinics* 29(4):39.

MacCara M. 1983. Extravasation: A hazard of intravenous therapy. *Drug Intelligence Clin Pharm* 17:713.

Further Information

Peter Glass provides a strong rationale for intravenous therapy including pharmacokinetic and pharma-codynamic bases for continuous delivery. Clinical complications around intravenous therapy are well summarized by MacCara [1983] and Bohony [1993]. The AAMI Standard for Infusion Devices provides a comprehensive means of evaluating infusion device technology, and the British Department of Health OHEI Report #198 provides a competitive analysis of pumps and controllers.

28
Essentials of Anesthesia Delivery

A.William Paulsen
Emory University

The intent of this chapter is to provide an introduction to the practice of anesthesiology and to the technology currently employed. Limitations on the length of this work and the enormous size of the topic require that this chapter rely on other elements within this handbook and other texts cited as general references for many of the details that inquisitive minds desire and deserve.

The practice of anesthesia includes more than just providing relief from pain. In fact, pain relief can be considered a secondary facet of the specialty. In actuality, the modern concept of the safe and efficacious delivery of anesthesia requires consideration of three fundamental tenets, which are ordered here by relative importance:

1. maintenance of vital organ function
2. relief of pain
3. maintenance of the "internal milieu"

The first, maintenance of vital organ function, is concerned with preventing damage to cells and organ systems that could result from inadequate supply of oxygen and other nutrients. The delivery of blood and cellular substrates is often referred to as perfusion of the cells or tissues. During the delivery of an anesthetic, the patient's "vital signs" are monitored in an attempt to prevent inadequate tissue perfusion. However, the surgery itself, the patient's existing pathophysiology, drugs given for the relief of pain, or even the management of blood pressure may compromise tissue perfusion. Why is adequate perfusion of tissues a higher priority than providing relief of pain for which anesthesia is named? A rather obvious extreme example is that without cerebral perfusion, or perfusion of the spinal cord, delivery of an anesthetic is not necessary. Damage to other organ systems may result in a range of complications from delaying the patient's recovery to diminishing their quality of life to premature death.

In other words, the primary purpose of anesthesia care is to maintain adequate delivery of required substrates to each organ and cell, which will hopefully preserve cellular function. The second principle of anesthesia is to relieve the pain caused by surgery. Chronic pain and suffering caused by many disease states is now managed by a relatively new subspecialty within anesthesia, called *Pain Management*.

The third principle of anesthesia is the maintenance of the internal environment of the body, for example, the regulation of electrolytes (sodium, potassium, chloride, magnesium, calcium, etc.), acid-base balance, and a host of supporting functions on which cellular function and organ system communications rest.

0-8493-1813-0/03/$0.00+$1.50

The person delivering anesthesia may be an anesthesiologist (physician specializing in anesthesiology), an anesthesiology physician assistant (a person trained in a medical school at the master's level to administer anesthesia as a member of the care team lead by an Anesthesiologist), or a nurse anesthetist (a nurse with intensive-care-unit experience that has additional training in anesthesia provided by advanced practice nursing programs). There are three major categories of anesthesia provided to patients: (1) general anesthesia; (2) conduction anesthesia; and (3) monitored anesthesia care. General anesthesia typically includes the intravenous injection of anesthetic drugs that render the patient unconscious and paralyze their skeletal muscles. Immediately following drug administration a plastic tube is inserted into the trachea and the patient is connected to an electropneumatic system to maintain ventilation of the lungs. A liquid anesthetic agent is vaporized and administered by inhalation, sometimes along with nitrous oxide, to maintain anesthesia for the surgical procedure. Often, other intravenous agents are used in conjunction with the inhalation agents to provide what is called a balanced anesthetic.

Conduction anesthesia refers to blocking the conduction of pain and possibly motor nerve impulses traveling along specific nerves or the spinal cord. Common forms of conduction anesthesia include spinal and epidural anesthesia, as well as specific nerve blocks, for example, axillary nerve blocks. In order to achieve a successful conduction anesthetic, local anesthetic agents such as lidocaine, are injected into the proximity of specific nerves to block the conduction of electrical impulses. In addition, sedation may be provided intravenously to keep the patient comfortable while he or she is lying still for the surgery.

Monitored anesthesia care refers to monitoring the patient's vital signs while administering sedatives and analgesics to keep the patient comfortable, and treating complications related to the surgical procedure. Typically, the surgeon administers topical or local anesthetics to alleviate the pain.

In order to provide the range of support required, from the paralyzed mechanically ventilated patient to the patient receiving monitored anesthesia care, a versatile anesthesia delivery system must be available to the anesthesia care team. Today's anesthesia delivery system is composed of six major elements:

1. The primary and secondary sources of gases (O_2, air, N_2O, vacuum, gas scavenging, and possibly CO_2 and helium).
2. The gas blending and vaporization system.
3. The breathing circuit (including methods for manual and mechanical ventilation).
4. The excess gas-scavenging system that minimizes potential pollution of the operating room by anesthetic gases.
5. Instruments and equipment to monitor the function of the anesthesia delivery system.
6. Patient monitoring instrumentation and equipment.

The traditional anesthesia machine incorporated elements 1, 2, 3, and more recently 4. The evolution to the anesthesia delivery system adds elements 5 and 6. In the text that follows, references to the "anesthesia machine" refer to the basic gas-delivery system and breathing circuit as contrasted with the "anesthesia delivery system," which includes the basic "anesthesia machine" and all monitoring instrumentation.

28.1 Gases Used during Anesthesia and their Sources

Most inhaled anesthetic agents are liquids that are vaporized in a device within the anesthesia delivery system. The vaporized agents are then blended with other breathing gases before flowing into the breathing circuit and being administered to the patient. The most commonly administered form of anesthesia is called a balanced general anesthetic, and is a combination of inhalation agent plus intravenous analgesic drugs. Intravenous drugs often require electromechanical devices to administer an appropriately controlled flow of drug to the patient.

Gases needed for the delivery of anesthesia are generally limited to oxygen (O_2), air, nitrous oxide (N_2O), and possibly helium (He) and carbon dioxide (CO_2). Vacuum and gas scavenging lines are also required. There need to be secondary sources of these gases in the event of primary failure or questionable contamination. Typically, primary sources are those supplied from a hospital distribution system at 345 kPa (50 psig) through gas columns or wall outlets. The secondary sources of gas are cylinders hung on yokes on the anesthesia delivery system.

Oxygen

Oxygen provides an essential metabolic substrate for all human cells, but it is not without dangerous side effects. Prolonged exposure to high concentrations of oxygen may result in toxic effects within the lungs that decrease diffusion of gas into and out of the blood, and the return to breathing air following prolonged exposure to elevated O_2 may result in a debilitating explosive blood vessel growth in infants. Oxygen is usually supplied to the hospital in liquid form (boiling point of −183°C), stored in cryogenic tanks, and supplied to the hospital piping system as a gas. The efficiency of liquid storage is obvious since 1 liter of liquid becomes 860 liters of gas at standard temperature and pressure. The secondary source of oxygen within an anesthesia delivery system is usually one or more E cylinders filled with gaseous oxygen at a pressure of 15.2 MPa (2200 psig).

Air (78% N_2, 21% O_2, 0.9% Ar, 0.1% Other Gases)

The primary use of air during anesthesia is as a diluent to decrease the inspired oxygen concentration. The typical primary source of medical air (there is an important distinction between “air” and “medical air” related to the quality and the requirements for periodic testing) is a special compressor that avoids hydrocarbon-based lubricants for purposes of medical air purity. Dryers are employed to rid the compressed air of water prior to distribution throughout the hospital. Medical facilities with limited need for medical air may use banks of H cylinders of dry medical air. A secondary source of air may be available on the anesthesia machine as an E cylinder containing dry gas at 15.2 MPa.

Nitrous Oxide

Nitrous oxide is a colorless, odorless, and nonirritating gas that does not support human life. Breathing more than 85% N_2O may be fatal. N_2O is not an anesthetic (except under hyperbaric conditions), rather it is an analgesic and an amnestic. There are many reasons for administering N_2O during the course of an anesthetic including: enhancing the speed of induction and emergence from anesthesia; decreasing the concentration requirements of potent inhalation anesthetics (i.e., halothane, isoflurane, etc.); and as an essential adjunct to narcotic analgesics. N_2O is supplied to anesthetizing locations from banks of H cylinders that are filled with 90% liquid at a pressure of 5.1 MPa (745 psig). Secondary supplies are available on the anesthesia machine in the form of E cylinders, again containing 90% liquid. Continual exposure to low levels of N_2O in the workplace has been implicated in a number of medical problems including spontaneous abortion, infertility, birth defects, cancer, liver and kidney disease, and others. Although there is no conclusive evidence to support most of these implications, there is a recognized need to scavenge all waste anesthetic gases and periodically sample N_2O levels in the workplace to maintain the lowest possible levels consistent with reasonable risk to the operating room personnel and cost to the institution [Dorsch and Dorsch, 1998]. Another gas with analgesic properties similar to N_2O is xenon, but its use is experimental, and its cost is prohibitive at this time.

TABLE 28.1 Physical Properties of Gases Used during Anesthesia

GAS	Molecular Wt.	Density (g/L)	Viscosity (cp)	Specific Heat (KJ/Kg°C)
Oxygen	31.999	1.326	0.0203	0.917
Nitrogen	28.013	1.161	0.0175	1.040
Air	28.975	1.200	0.0181	1.010
Nitrous oxide	44.013	1.836	0.0144	0.839
Carbon dioxide	44.01	1.835	0.0148	0.850
Helium	4.003	0.1657	0.0194	5.190

TABLE 28.2 Physical Properties of Currently Available Volatile Anesthetic Agents

Agent Generic Name	Boiling Point (°C at 760 mmHg)	Vapor Pressure (mmHg at 20°C)	Liquid Density (g/ml)	MAC* (%)
Halothane	50.2	243	1.86	0.75
Enflurane	56.5	175	1.517	1.68
Isoflurane	48.5	238	1.496	1.15
Desflurane	23.5	664	1.45	6.0
Sevoflurane	58.5	160	1.51	2.0

[*] **M**inimum **A**lveolar **C**oncentration is the percent of the agent required to provide surgical anesthesia to 50% of the population in terms of a cummulative dose response curve. The lower the MAC, the more potent the agent.

Carbon Dioxide

Carbon dioxide is colorless and odorless, but very irritating to breathe in higher concentrations. CO_2 is a byproduct of human cellular metabolism and is not a life-sustaining gas. CO_2 influences many physiologic processes either directly or through the action of hydrogen ions by the reaction $CO_2+H_2O\leftrightarrow H_2CO_3\leftrightarrow H^++HCO_3^-$. Although not very common in the U.S. today, in the past CO2 was admini$_{st}$ered during anesthesia to stimulate respiration that was depressed by anesthetic agents and to cause increased blood flow in otherwise compromised vasculature during some surgical procedures. Like N_2O, CO_2 is supplied as a liquid in H cylinders for distribution in pipeline systems or as a liquid in E cylinders that are located on the anesthesia machine.

Helium

Helium is a colorless, odorless, and nonirritating gas that will not support life. The primary use of helium in anesthesia is to enhance gas flow through small orifices as in asthma, airway trauma, or tracheal stenosis. The viscosity of helium is not different from other anesthetic gases (refer to Table 28.1) and is therefore of no benefit when airway flow is laminar. However, in the event that ventilation must be performed through abnormally narrow orifices or tubes that create turbulent flow conditions, helium is the preferred carrier gas. Resistance to turbulent flow is proportional to the density rather than viscosity of the gas and helium is an order of magnitude less dense than other gases. A secondary advantage of helium is that it has a large specific heat relative to other anesthetic gases and therefore can carry the heat from laser surgery out of the airway more effectively than air, oxygen, or nitrous oxide.

28.2 Gas Blending and Vaporization System

The basic anesthesia machine utilizes primary low-pressure gas sources of 345 kPa (50 psig) available from wall or ceiling column outlets, and secondary high pressure gas sources located on the machine as pictured schematically in Fig. 28.1. Tracing the path of oxygen in the machine demonstrates that oxygen comes from either the low-pressure source, or from the 15.2 Mpa (2200 psig), high-pressure yokes via

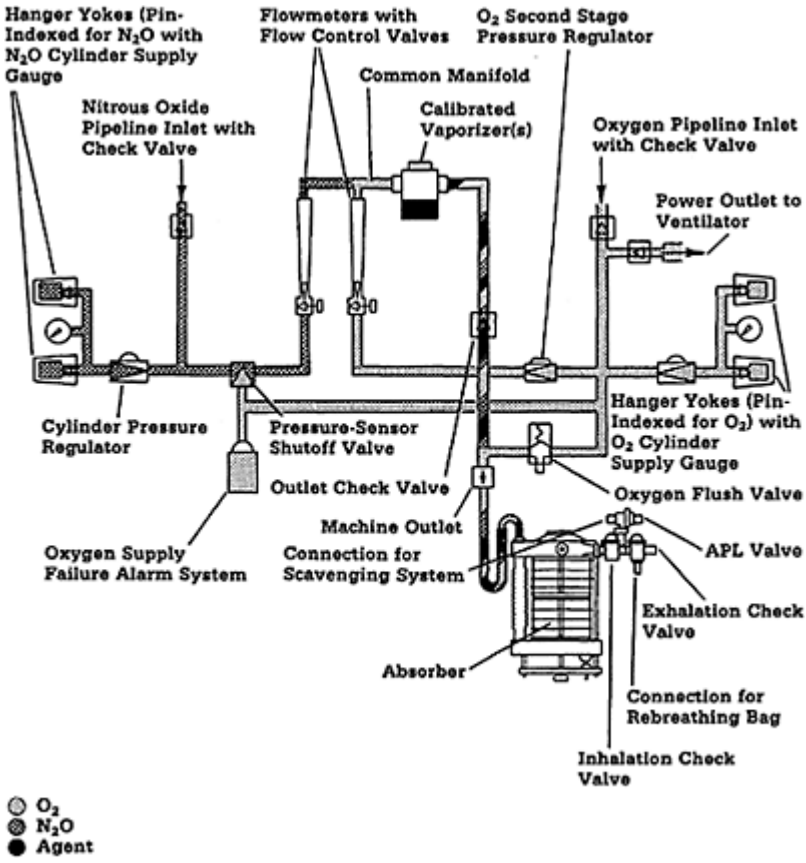


FIGURE 28.1 Schematic diagram of gas piping within a simple two-gas (oxygen and nitrous oxide) anesthesia machine.

cylinder pressure regulators and then branches to service several other functions. First and foremost, the second-stage pressure regulator drops the O_2 pressure to approximately 110 kPa (16 psig) before it enters the needle valve and the rotameter type flowmeter. From the flowmeter O_2 mixes with gases from other flowmeters and passes through a calibrated agent vaporizer, where specific inhalation anesthetic agents are vaporized and added to the breathing gas mixture. Oxygen is also used to supply a reservoir canister that sounds a reed alarm in the event that the oxygen pressure drops below 172 kPa (25 psig). When the oxygen pressure drops to 172 kPa or lower, then the nitrous oxide pressure sensor shutoff valve closes and N_2O is prevented from entering its needle valve and flowmeter and is therefore eliminated from the breathing gas mixture. In fact, all machines built in the U.S. have pressure-sensor shutoff valves installed in the lines to every flowmeter, except oxygen, to prevent the delivery of a hypoxic gas mixture in the

event of an oxygen pressure failure. Oxygen may also be delivered to the common gas outlet or machine outlet via a momentary normally closed flush valve that typically provides a flow of 65 to 80 liters of O_2 per minute directly into the breathing circuit. Newer machines are required to have a safety system for limiting the minimum concentration of oxygen that can be delivered to the patient to 25%. The flow paths for nitrous oxide and other gases are much simpler in the sense that after coming from the high-pressure regulator or the low-pressure hospital source, gas is immediately presented to the pressure-sensor shutoff valve from where it travels to its specific needle valve and flowmeter to join the common gas line and enter the breathing circuit.

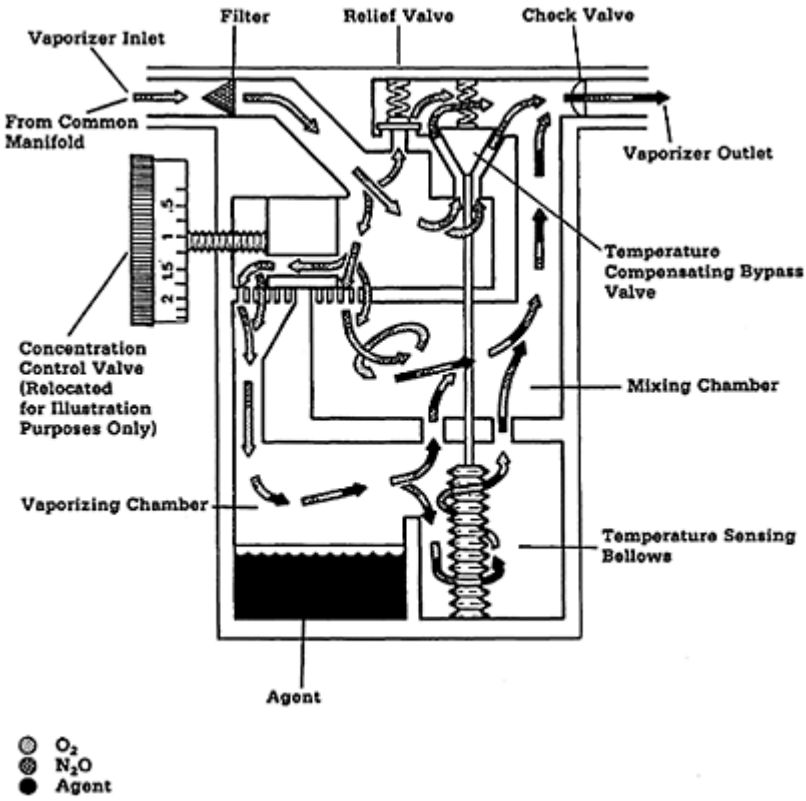


FIGURE 28.2 Schematic diagram of a calibrated in-line vaporizer that uses the flow-over technique for adding anesthetic vapor to the breathing gas mixture.

Currently all anesthesia machines manufactured in the U.S. use only calibrated flowthrough vaporizers, meaning that all of the gases from the various flowmeters are mixed in the manifold prior to entering the vaporizer. Any given vaporizer has a

calibrated control knob that, once set to the desired concentration for a specific agent, will deliver that concentration to the patient. Some form of interlock system must be provided such that only one vaporizer may be activated at any given time. Figure 28.2 schematically illustrates the operation of a purely mechanical vaporizer with temperature compensation. This simple flowover design permits a fraction of the total gas flow to pass into the vaporizing chamber where it becomes saturated with vapor before being added back to the total gas flow. Mathematically this is approximated by:

$$F_A = \frac{Q_{VC} * P_A}{P_B * \left(Q_{VC} + Q_G\right) - P_A * Q_G},$$

where F_A is the fractional concentration of agent at the outlet of the vaporizer, Q_G is the total flow of gas entering the vaporizer, Q_{vc} is the amount of Q_G that is diverted into the vaporization chamber, P_A is the vapor pressure of the agent, and P_B is the barometric pressure.

From Fig. 28.2, the temperature compensator would decrease Q_{vc} as temperature increased because vapor pressure is proportional to temperature. The concentration accuracy over a range of clinically expected gas flows and temperatures is approximately ±15%. Since vaporization is an endothermic process, anesthetic vaporizers must have sufficient thermal mass and conductivity to permit the vaporization process to proceed independent of the rate at which the agent is being used.

28.3 Breathing Circuits

The concept behind an effective breathing circuit is to provide an adequate volume of a controlled concentration of gas to the patient during inspiration, and to carry the exhaled gases away from the patient during exhalation. There are several forms of breathing circuits that can be classified into 2 basic types; (1) open circuit, meaning no rebreathing of any gases and no CO_2 absorber present; and (2) closed circuit, indicating presence of CO_2 absorber and some rebreathing of other gases. Figure 28.3 illustrates the Lack modification of a Mapleson open-circuit breathing system. There are no valves and no CO_2 absorber. There is a great potential for the patient to rebreath their own exhaled gases unless the fresh gas inflow is 2 to 3 times the patient's minute volume. Figure 28.4 illustrates the most popular form of breathing circuit, the circle system, with oxygen monitor, circle pressure gage, volume monitor (spirometer), and airway pressure sensor. The circle is a closed system, or semiclosed when the fresh gas inflow exceeds the patient's requirements. Excess gas evolves into the scavenging device, and some of the exhaled gas is rebreathed after having the CO_2 removed. The inspiratory and expiratory valves in the circle system guarantee that gas flows to the patient from the inspiratory limb and away from the patient through the

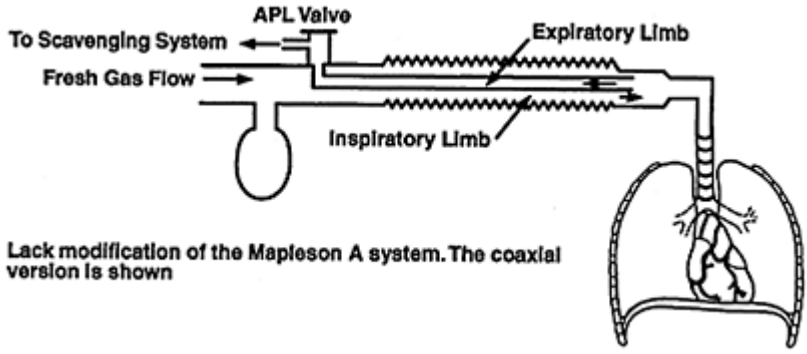


FIGURE 28.3 An example of an open-circuit breathing system that does not use unidirectional flow valves or contain a carbon dioxide absorbent.

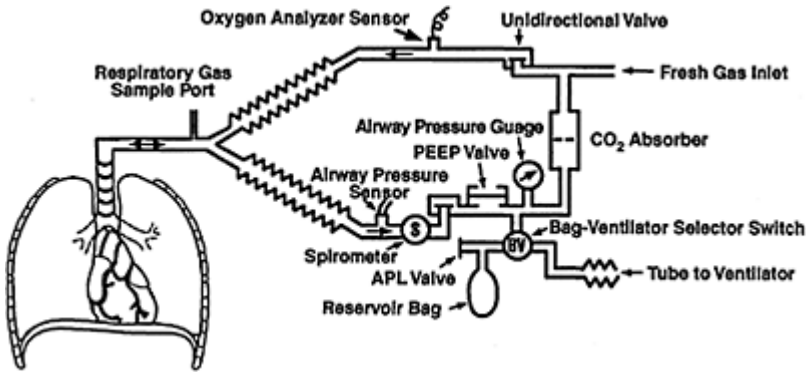


FIGURE 28.4 A diagram of a closed-circuit circle breathing system with unidirectional valves, inspired oxygen sensor, pressure sensor, and CO_2 absorber.

exhalation limb. In the event of a failure of either or both of these valves, the patient will rebreathe exhaled gas that contains CO_2, which is a potentially dangerous situation.

There are two forms of mechanical ventilation used during anesthesia: (1) volume ventilation, where the volume of gas delivered to the patient remains constant regardless of the pressure that is required; and (2) pressure ventilation, where the ventilator provides whatever volume to the patient that is required to produce some desired pressure in the breathing circuit. Volume ventilation is the most popular since the volume delivered remains theoretically constant despite changes in lung compliance. Pressure ventilation is

useful when compliance losses in the breathing circuit are high relative to the volume delivered to the lungs.

Humidification is an important adjunct to the breathing circuit because it maintains the integrity of the cilia that line the airways and promote the removal of mucus and particulate matter from the lungs. Humidification of dry breathing gases can be accomplished by simple passive heat and moisture exchangers inserted into the breathing circuit at the level of the endotracheal tube connectors, or by elegant dual servo-electronic humidifiers that heat a reservoir rilled with water and also heat a wire in the gas delivery tube to prevent rain-out of the water before it reaches the patient. Electronic safety measures must be included in these active devices due to the potential for burning the patient and the fire hazard.

28.4 Gas Scavenging Systems

The purpose of scavenging exhaled and excess anesthetic agents is to reduce or eliminate the potential hazard to employees who work in the environment where anesthetics are administered, including operating rooms, obstetrical areas, special procedures areas, physician's offices, dentist's offices, and veterinarian's surgical suites. Typically more gas is administered to the breathing circuit than is required by the patient, resulting in the necessity to remove excess gas from the circuit. The scavenging system must be capable of collecting gas from all components of the breathing circuit, including adjustable pressure level valves, ventilators, and sample withdrawal type gas monitors, without altering characteristics of the circuit such as pressure or gas flow to the patient. There are two broad types of scavenging systems as illustrated in Fig. 28.5: the open interface is a simple design that requires a large physical space for the reservoir volume, and the closed interface with an expandable reservoir bag and that must include relief valves for handling the cases of no scavenged flow and great excess of scavenged flow.

Trace gas analysis must be performed to guarantee the efficacy of the scavenging system. The National Institutes of Occupational Safety and Health (NIOSH) recommends that trace levels of nitrous oxide be maintained at or below 25 parts per million (ppm) time weighted average and that halogenated anesthetic agents remain below 2 ppm.

28.5 Monitoring the Function of the Anesthesia Delivery System

The anesthesia machine can produce a single or combination of catastrophic events, any one of which could be fatal to the patient:

1. delivery of a hypoxic gas mixture to the patient;
2. the inability to adequately ventilate the lungs by not producing positive pressure in the patient's lungs, by not delivering an adequate volume of gas to the lungs, or by improper breathing circuit connections that permit the patient's lungs to receive only rebreathed gases;

3. the delivery of an overdose of an inhalational anesthetic agent.

The necessary monitoring equipment to guarantee proper function of the anesthesia delivery system includes at least:

• Inspired oxygen concentration monitor with absolute low level alarm of 19%.
• Airway pressure monitor with alarms for:

1. low pressure indicative of inadequate breathing volume and possible leaks
2. sustained elevated pressures that could compromise cardiovascular function

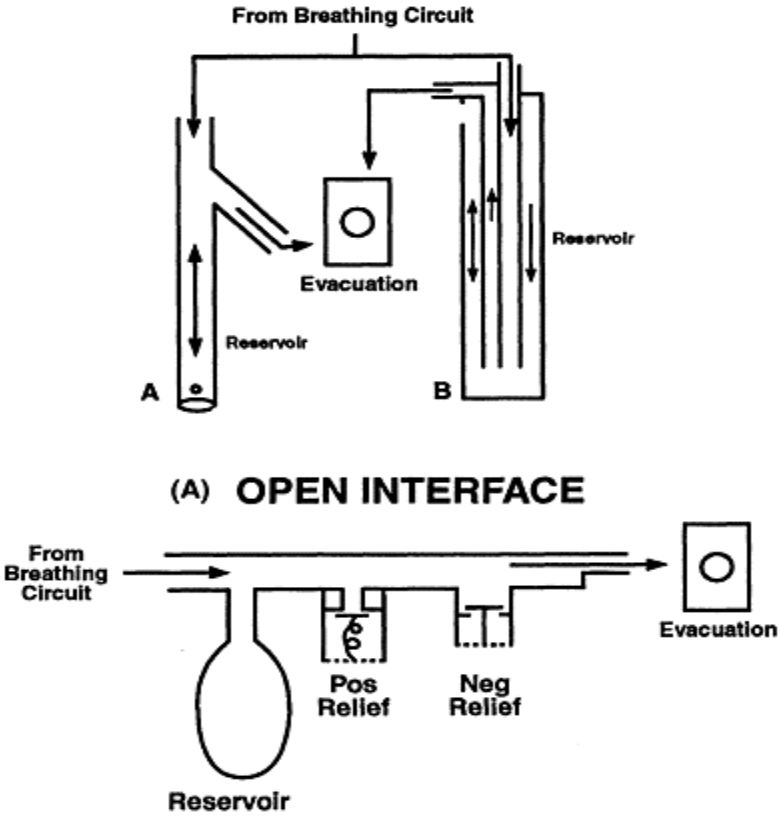


FIGURE 28.5 Examples of open and closed gas scavenger interfaces. The closed interface requires relief valves in the event of scavenging flow failure.

3.high pressures that could cause pulmonary barotrauma
4.subatmospheric pressure that could cause collapse of the lungs

• Exhaled gas volume monitor.
• Carbon dioxide monitor (capnography).
• Inspired and exhaled concentration of anesthetic agents by any of the following:

1. mass spectrometer
2. Raman spectrometer
3. infrared or other optical spectrometer

A mass spectrometer is a very useful cost-effective device since it alone can provide capnography, inspired and exhaled concentrations of all anesthetic agents, plus all breathing gases simultaneously (O_2, N_2, CO_2, N_2O, Ar, He, halothane, enflurane, isoflurane, desflurane, and suprane). The mass spectrometer is unique in that it may be tuned to monitor an assortment of exhaled gases while the patient is asleep, including: (1) ketones for detection of diabetic ketoacidosis; (2) ethanol or other marker in the irrigation solution during transurethral resection of the prostate for early detection of the TURP syndrome, which results in a severe dilution of blood electrolytes; and (3) pentanes during the evolution of a heart attack, to mention a few.

Sound monitoring principles require: (1) earliest possible detection of untoward events (before they result in physiologic derangements); and (2) specificity that results in rapid identification and resolution of the problem. An extremely useful rule to always consider is ***"never monitor the anesthesia delivery system performance through the patient's physiologic responses."*** That is, never intentionally use a device like a pulse oximeter to detect a breathing circuit disconnection since the warning is very late and there is no specific information provided that leads to rapid resolution of the problem.

28.6 Monitoring the Patient

The anesthetist's responsibilities to the patient include: providing relief from pain and preserving all existing normal cellular function of all organ systems. Currently the latter obligation is fulfilled by monitoring essential physiologic parameters and correcting any substantial derangements that occur before they are translated into permanent cellular damage. The inadequacy of current monitoring methods can be appreciated by realizing that most monitoring modalities only indicate damage after an insult has occurred, at which point the hope is that it is reversible or that further damage can be prevented.

Standards for basic intraoperative monitoring of patients undergoing anesthesia, which were developed and adopted by the American Society of Anesthesiologists, became effective in 1990. Standard I concerns the responsibilities of anesthesia personnel, while Standard II requires that the patient's oxygenation, ventilation, circulation, and temperature be evaluated continually during all anesthetics. The following list indicates the instrumentation typically available during the administration of anesthetics.

Electrocardiogram	Noninvasive or invasive blood pressure
Pulse oximetry	Temperature

Urine output	Nerve stimulators
Cardiac output	Mixed venous oxygen saturation
Electroencephalogram (EEC)	Transesophageal echo-cardiography (TEE)
Evoked potentials	Coagulation status

Blood gases and electrolytes (Po_2, Pco_2, pH, BE, Na^+, K^+, Cl^-, Ca^{++}, and glucose)

Mass spectrometry, raman spectrometry, or infrared breathing gas analysis

Control of Patient Temperature

Anesthesia alters the thresholds for temperature regulation and the patient becomes unable to maintain normal body temperature. As the patient's temperature falls even a few degrees toward room temperature, several physiologic derangements occur: (1) drug action is prolonged; (2) blood coagulation is impaired; and (3) postoperative infection rate increases. On the positive side, cerebral protection from inadequate perfusion is enhanced by just a few degrees of cooling. Proper monitoring of core body temperature and forced hot-air warming of the patient is essential.

Monitoring the Depth of Anesthesia

There are two very unpleasant experiences that patients may have while undergoing an inadequate anesthetic: (1) the patient is paralyzed and unable to communicate their state of discomfort, and they are feeling the pain of surgery and are aware of their surroundings; (2) the patient may be paralyzed, unable to communicate, and is aware of their surroundings, but is not feeling any pain. The ability to monitor the depth of anesthesia would provide a safeguard against these unpleasant experiences. However, despite numerous instruments and approaches to the problem it remains elusive. Brainstem auditory evoked responses have come the closest to depth-of-anesthesia monitoring, but it is difficult to perform, is expensive, and is not possible to perform during many types of surgery. A promising new technology, called bispectral index (BIS monitoring) is purported to measure the level of patient awareness through multivariate analysis of a single channel of the EEC.

Anesthesia Computer-Aided Record Keeping

Conceptually, every anesthetist desires an automated anesthesia record keeping system. Anesthesia care can be improved through the feedback provided by correct record keeping, but today's systems have an enormous overhead associated with their use when compared to standard paper record keeping. No doubt that automated anesthesia record keeping reduces the drudgery of routine recording of vital signs, but to enter drugs and drips and their dosages, fluids administered, urine output, blood loss, and other data requires much more time and machine interaction than the current paper system. Despite attempts to use every input/output device ever produced by the computer industry from keyboards to bar codes to voice and handwriting recognition, no solution has been found that meets wide acceptance. Tenets of a successful system must include:

1. The concept of a user transparent system, which is ideally defined as requiring no communication between the computer and the clinician (far beyond the concept of user friendly), and therefore that is intuitively obvious to use even to the most casual users.
2. Recognition of the fact that educational institutions have very different requirements from private-practice institutions.
3. Real-time hard copy of the record produced at the site of anesthetic administration that permits real-time editing and notation.
4. Ability to interface with a great variety of patient and anesthesia delivery system monitors from various suppliers.
5. Ability to interface with a large number of hospital information systems.
6. Inexpensive to purchase and maintain.

Alarms

Vigilance is the key to effective risk management, but maintaining a vigilant state is not easy. The practice of anesthesia has been described as moments of sheer terror connected by times of intense boredom. Alarms can play a significant role in redirecting one's attention during the boredom to the most important event regarding patient safety, but only if false alarms can be eliminated, alarms can be prioritized, and all alarms concerning anesthetic management can be displayed in a single clearly visible location.

Ergonomics

The study of ergonomics attempts to improve performance by optimizing the relationship between people and their work environment. Ergonomics has been defined as a discipline that investigates and applies information about human requirements, characteristics, abilities, and limitations to the design, development, and testing of equipment, systems, and jobs [Loeb, 1993]. This field of study is only in its infancy and examples of poor ergonomic design abound in the anesthesia workplace.

Simulation in Anesthesia

Complete patient simulators are hands-on realistic simulators that interface with physiologic monitoring equipment to simulate patient responses to equipment malfunctions, operator errors, and drug therapies. There are also crisis management simulators. Complex patient simulators, which are analogous to flight simulators, are currently being marketed for training anesthesia personnel. The intended use for these complex simulators is currently being debated in the sense that training people to respond in a preprogrammed way to a given event may not be adequate training.

Reliability

The design of an anesthesia delivery system is unlike the design of most other medical devices because it is a life-support system. As such, its core elements deserve all of the considerations of the latest failsafe technologies. Too often in today's quest to apply

microprocessor technology to everything, tradeoffs are made among reliability, cost, and engineering elegance. The most widely accepted anesthesia machine designs continue to be based upon simple ultrareliable mechanical systems with an absolute minimum of catastrophic failure modes. The replacement of needle valves and rotameters, for example, with microprocessor controlled electromechanical valves can only introduce new catastrophic failure modes. However, the inclusion of microprocessors can enhance the safety of anesthesia delivery if they are implemented without adding catastrophic failure modes.

Further Information

Blitt, C.D. and Hines R.L., Eds. 1995. *Monitoring in Anesthesia and Critical Care Medicine,* 3rd ed. Churchill Livingstone, New York.

Dorsch, J.A. and Dorsch S.E. 1998. *Understanding Anesthesia Equipment,* 4th ed. Williams and Wilkins, Baltimore, MD.

Ehrenwerth, J. and Eisenkraft, J.B. 1993. *Anesthesia Equipment: Principles and Applications.* Mosby, St. Louis, MO.

Gravenstein N. and Kirby R.R., Eds. 1996. *Complications in Anesthesiology,* 2nd ed. Lippincott-Raven, Philadelphia, PA.

Loeb, R. 1993. Ergonomics of the anesthesia workplace. *STA Interface* 4(3):18.

Miller, R.D. Ed. 1999. *Anesthesia,* 5th ed. Churchill Livingstone, New York.

Miller, R.D. Ed. 1998. *Atlas of Anesthesia.* Churchill Livingstone, New York.

Saidman, LJ. and Smith, N.T., Eds. 1993. *Monitoring in Anesthesia,* 3rd ed. Butterworth-Heinemann, Stoneham, MA.

29
Biomedical Lasers

Millard M.Judy
Baylor Research Institute

Approximately 20 years ago the CO_2 laser was introduced into surgical practice as a tool to photo thermally ablate, and thus to incise and to debulk, soft tissues. Subsequently, three important factors have led to the expanding biomedical use of laser technology, particularly in surgery. These factors are: (1) the increasing understanding of the wavelength selective interaction and associated effects of *ultraviolet-infrared (UV-IR) radiation* with biologic tissues, including those of acute damage and long-term healing, (2) the rapidly increasing availability of lasers emitting (essentially monochromatically) at those wavelengths that are strongly absorbed by molecular species within tissues, and (3) the availability of both optical fiber and lens technologies as well as of endoscopic technologies for delivery of the laser radiation to the often remote internal treatment site. Fusion of these factors has led to the development of currently available biomedical laser systems.

This chapter briefly reviews the current status of each of these three factors. In doing so, each of the following topics will be briefly discussed:

1. The physics of the interaction and the associated effects (including clinical efforts) of UV-IR radiation on biologic tissues.
2. The fundamental principles that underlie the operation and construction of all lasers.
3. The physical properties of the optical delivery systems used with the different biomedical lasers for delivery of the laser beam to the treatment site.
4. The essential physical features of those biomedical lasers currently in routine use ranging over a number of clinical specialties, and brief descriptions of their use.
5. The biomedical uses of other lasers used surgically in limited scale or that are currently being researched for applications in surgical and diagnostic procedures and the photosensitized inactivation of cancer tumors.

In this review, effort is made in the text and in the last section to provide a number of key references and sources of information for each topic that will enable the reader's more in-depth pursuit.

0-8493-1813-0/03/$0.00+$1.50

29.1 Interaction and Effects of UV-IR Laser Radiation on Biologic Tissues

Electromagnetic radiation in the UV-IR spectral range propagates within biologic tissues until it is either scattered or absorbed.

Scattering in Biologic Tissue

Scattering in matter occurs only at the boundaries between regions having different optical refractive indices and is a process in which the energy of the radiation is conserved [Van de Hulst, 1957]. Since biologic tissue is structurally inhomogeneous at the microscopic scale, e.g., both subcellular and cellular dimensions, and at the macroscopic scale, e.g., cellular assembly (tissue) dimensions, and predominantly contains water, proteins, and lipids, all different chemical species, it is generally regarded as a scatterer of UV-IR radiation. The general result of scattering is deviation of the direction of propagation of radiation. The deviation is strongest when wavelength and scatterer are comparable in dimension (Mie scattering) and when wavelength greatly exceeds particle size (Rayleigh scattering) [Van de Hulst, 1957]. This dimensional relationship results in the deeper penetration into biologic tissues of those longer wavelengths that are not absorbed appreciably by pigments in the tissues. This results in the relative transparency of nonpigmented tissues over the visible and near-IR wavelength ranges.

Absorption in Biologic Tissue

Absorption of UV-IR radiation in matter arises from the wavelength-dependent resonant absorption of radiation by molecular electrons of optically absorbing molecular species [Grossweiner, 1989]. Because of the chemical inhomogeneity of biologic tissues, the degree of absorption of incident radiation strongly depends upon its wavelength. The most prevalent or concentrated UV-IR absorbing molecular species in biologic tissues are listed in Table 29.1 along with associated high-absorbance wavelengths. These species include the peptide bonds; the phenylalanine, tyrosine, and tryptophan residues of proteins, all of which absorb in the UV range; oxy- and deoxyhemoglobin of blood which absorb in the visible to near-IR range; melanin, which absorbs throughout the UV to near-IR range, that decreasing absorption occurring with increasing wavelength; and water, which absorbs maximally in the mid-IR range [Hale and Querry, 1973; Miller and Veitch, 1993; White et al., 1968]. Biomedical lasers and their emitted radiation wavelength values also are tabulated also in Table 29.1. The correlation between the wavelengths of clinically useful lasers and wavelength regions of absorption by constituents of biological tissues is evident. Additionally, exogenous light-absorbing chemical species may be intentionally present in tissues. These include:

1. Photosensitizers, such as porphyrins, which upon excitation with UV-visible light initiate photo-chemical reactions that are cytotoxic to the cells of the tissue, e.g., a cancer that concentrates the photosensitizer relative to surrounding tissues [Spikes, 1989].

2. Dyes such as indocyanine green which, when dispersed in a concentrate fibrin protein gel can be used to localize 810 nm *GaAlAs* diode laser radiation and the associated heating to achieve localized thermal denaturation and bonding of collagen to effect joining or welding of tissue [Bass et al., 1992; Oz et al., 1989].
3. Tattoo pigments including graphite (black) and black, blue, green, and red organic dyes [Fitz-patrick, 1994; McGillis et al., 1994].

TABLE 29.1 UV-IR-Radiation-Absorbing Constituent of Biological Tissues and Biomedical Laser Wavelengths

Constituent	Tissue Type	Optical Absorption		Laser Type	Wavelength, nm
		Wavelength,* nm	Relative† Strength		
Proteins	All				
Peptide bond		<220 (r)	+++++++	ArF	193
Amino acid					
Residues					
Tryptophan		220–290 (r)	+		
Tyrosine		220–290 (r)	+		
Phenylalanine		220–2650 (r)	+		
Pigments					
Oxyhemoglobin	Blood	414 (p)	+++	Ar ion	488–514.5
	vascular	537 (p)	++	Frequency	532
	tissues	575 (p)	++	Doubled	
		970 (p)	+	Nd: YAG	
		(690–1100) (r)		Diode	810
Deoxyhemoglobin	Blood	431 (p)	+++	Nd: YAG Dye	1064 400–700
	vascular	554 (p)	++	Nd: YAG	1064
	tissues				
Melanin	Skin	220–1000 (r)	++++	Ruby	693
Water	All	2–1 (p)	+++	Ho: YAG	2100
		3.02 (p)	+++++++	Er: YAG	2940
		>2.94 (r)	++++	CO_2	10,640

* (p): Peak absorption wavelength; (r): wavelength range.
† The number of + signs qualitatively ranks the magnitude of the optical absorbtion.

29.2 Penetration and Effects of UV-IR Laser Radiation into Biologic Tissue

Both scattering and absorption processes affect the variations of the intensity of radiation with propagation into tissues. In the absence of scattering, absorption results in an exponential decrease of radiation intensity described simply by Beers law [Grossweiner, 1989]. With appreciable scattering present, the decrease in incident intensity from the surface is no longer monotonic. A maximum in local internal intensity is found to be present due to efficient backscattering, which adds to the intensity of the incoming beam as shown, for example, by Miller and Veitch [1993] for visible light penetrating into the skin and by Rastegar and coworkers [1992] for 1.064 μm *Nd: YAG* laser radiation penetrating into the prostate gland. Thus, the relative contributions of absorption and scattering of incident laser radiation will stipulate the depth in a tissue at which the resulting tissue effects will be present. Since the absorbed energy can be released in a number of different ways including thermal vibrations, fluorescence, and resonant electronic energy transfer according to the identity of the absorber, the effects on tissue are in general different. Energy release from both hemoglobin and melanin pigments and from water is by molecular vibrations resulting in a local temperature rise. Sufficient continued energy absorption and release can result in local temperature increases which, as energy input increases, result in protein denaturation (41 to 65°C), water evaporation and boiling (up to ≈300°C under confining pressure of tissue), thermolysis of proteins, generation of gaseous decomposition products and of carbonaceous residue or char (≥300°C). The generation of residual char is minimized by sufficiently rapid energy input to support rapid gasification reactions. The clinical effect of this chain of thermal events is tissue ablation. Much smaller values of energy input result in coagulation of tissues due to protein denaturation.

Energy release from excited exogenous photosensitizing dyes is via formation of free-radical species or energy exchange with itinerant dissolved molecular oxygen [Spikes, 1989]. Subsequent chemical reactions following free-radical formation or formation of an activated or more reactive form of molecular oxygen following energy exchange can be toxic to cells with takeup of the photosensitizes

Energy release following absorption of *visible (VIS) radiation* by fluorescent molecular species, either endogenous to tissue or exogenous, is predominantly by emission of longer wavelength radiation [Lakowicz, 1983]. Endogenous fluorescent species include tryptophan, tyrosine, phenylalanine, flavins, and metal-free porphyrins. Comparison of measured values of the intensity of fluorescence emission from hyperplastic (transformed precancerous) cervical cells to cancerous cervical cells with normal cervical epithelial cells shows a strong potential for diagnostic use in the automated diagnosis and staging of cervical cancer [Mahadevan et al., 1993].

29.3 Effects of Mid-IR Laser Radiation

Because of the very large absorption by water of radiation with wavelength in the IR range ≥2.0 μm, the radiation of *Ho: YAG*, *Er: YAG*, and CO_2 lasers is absorbed within a

very short distance of the tissue surface, and scattering is essentially unimportant. Using published values of the water absorption coefficient [Hale and Querry, 1973] and assuming an 80% water content and that the decrease in intensity is exponential with distance, the depth in the "average" soft tissue at which the intensity has decreased to 10% of the incident value (the optical penetration depth) is estimated to be 619, 13, and 170 micrometers, respectively, for Ho: YAG, Er: YAG, and CO_2 laser radiation. Thus, the absorption of radiation from these laser sources and thermalization of this energy results essentially in the formation of a surface heat source. With sufficient energy input, tissue ablation through water boiling and tissue thermolysis occur at the surface. Penetration of heat to underlying tissues is by diffusion alone; thus, the depth of coagulation of tissue below the surface region of ablation is limited by competition between thermal diffusion and the rate of descent of the heated surface impacted by laser radiation during ablation of tissue. Because of this competition, coagulation depths obtained in soft biologic tissues with use of mid-IR laser radiation are typically ≤205 to 500 μm, and the ability to achieve sealing of blood vessels leading to hemostatic ("bloodless") surgery is limited [Judy et al., 1992; Schroder et al., 1987].

29.4 Effects of Near-IR Laser Radiation

The 810-nm and 1064-μm radiation, respectively, of the GaAlAs diode laser and Nd: YAG laser penetrate more deeply into biologic tissues than the radiation of longer-wavelength IR lasers. Thus, the resulting thermal effects arise from absorption at greater depth within tissues, and the depths of coagulation and degree of hemostasis achieved with these lasers tend to be greater than with the longer-wavelength IR lasers. For example, the optical penetration depths (10% incident intensity) for 810-nm and 1.024-μm radiation are estimated to be ϕ4.6 and ϕ8.6 mm, respectively, in canine prostate tissue [Rastegar et al., 1992]. Energy deposition of 3600 J from each laser onto the urethral surface of the canine prostate results in maximum coagulation depths of ϕ8 and 12 mm, respectively, using diode and Nd: YAG lasers [Motamedi et al., 1993]. Depths of optical penetration and coagulation in porcine liver, a more vascular tissue than prostate gland, of ϕ2.8 and 9.6 mm, respectively, were obtained with a Nd: YAG laser beam, and of 7 and 12 mm, respectively, with an 810-nm diode laser beam [Rastegar et al., 1992]. The smaller penetration depth obtained with 810-nm diode radiation in liver than in prostate gland reflects the effect of greater vascularity (blood content) on near-IR propagation.

29.5 Effects of Visible-Range Laser Radiation

Blood and vascular tissues very efficiently absorb radiation in the visible wavelength range due to the strong absorption of hemoglobin. This absorption underlies, for example, the use of:

1. The argon ion laser (488 to 514.5 nm) in the localized heating and thermal coagulation of the vascular choroid layer and adjacent retina, resulting in the anchoring of the retina in treatment of retinal detachment [Katoh and Peyman, 1988].
2. The argon ion laser (488 to 514.5 nm), frequency-doubled Nd: YAG laser (532 nm), and dye laser radiation (585 nm) in the coagulative treatment of cutaneous vascular lesions such as port wine stains [Mordon et al., 1993].
3. The argon ion (488 to 514.5 nm) and frequency-doubled Nd: YAG lasers (532 nm) in the ablation of pelvic endometrial lesions which contain brown iron-containing pigments [Keye et al., 1983].

Because of the large absorption by hemoglobin and iron-containing pigments, the incident laser radiation is essentially absorbed at the surface of the blood vessel or lesion, and the resulting thermal effects are essentially local [Miller and Veitch, 1993].

29.6 Effects of UV Laser Radiation

Whereas exposure of tissue to IR and visible-light-range laser energy result in removal of tissue by thermal ablation, exposure to *argon fluoride (ArF)* laser radiation of 193-nm wavelength results predominantly in ablation of tissue initiated by a photochemical process [Garrison and Srinivasan, 1985]. This ablation arises from repulsive forces between like-charged regions of ionized protein molecules that result from ejection of molecular electrons following UV photon absorption [Garrison and Srinivasan, 1985]. Because the ionization and repulsive processes are extremely efficient, little of the incident laser energy escapes as thermal vibrational energy, and the extent of thermal coagulation damage adjacent to the site of incidence is very limited [Garrison and Srinivasan, 1985]. This feature and the ability to tune very finely the fluence emitted by the ArF laser so that micrometer depths of tissue can be removed have led to ongoing clinical trials to investigate the efficiency of the use of the ArF laser to selectively remove tissue from the surface of the human cornea for correction of short-sighted vision to eliminate the need for corrective eyewear [Van Saarloos and Constable, 1993].

29.7 Effects of Continuous and Pulsed IR-Visible Laser Radiation and Association Temperature Rise

Heating following absorption of IR-visible laser radiation arises from molecular vibration during loss of the excitation energy and initially is manifested locally within the exposed region of tissue. If incidence of the laser energy is maintained for a sufficiently long time, the temperature within adjacent regions of biologic tissue increases due to heat diffusion. The mean squared distance $\langle X^2 \rangle$ over which appreciable heat diffusion and temperature rise occur during exposure time t can be described in terms of the thermal diffusion time τ by the equation:

$$\left\langle X^2 \right\rangle = \tau t \,, \tag{29.1}$$

where τ is defined as the ratio of the thermal conductivity to the product of the heat capacity and density. For soft biologic tissues τ is approximately 1×10^{3} cm^2 s^{-1} [Meijering et al., 1993]. Thus, with continued energy input, the distance over which thermal diffusion and temperature rise occurs increases. Conversely, with use of pulsed radiation, the distance of heat diffusion can be made very small; for example, with exposure to a 1-μs pulse, the mean thermal diffusion distance is found to be approximately 0.3 μm, or about 3 to 10% of a biologic cell diameter. If the laser radiation is strongly absorbed and the ablation of tissues is efficient, then little energy diffuses away from the site of incidence, and lateral thermally induced coagulation of tissue can be minimized with pulses of short duration. The effect of limiting lateral thermal damage is desirable in the cutting of the cornea [Hibst et al., 1992] and sclera of the eye [Hill et al., 1993], and joint cartilage [Maes and Sherk, 1994], all of which are avascular (or nearly so, with cartilage), and the hemostasis arising from lateral tissue coagulation is not required.

29.8 General Description and Operation of Lasers

Lasers emit a beam of intense electromagnetic radiation that is essentially monochromatic or contains at most a few nearly monochromatic wavelengths and is typically only weakly divergent and easily focused into external optical systems. These attributes of laser radiation depend on the key phenomenon that underlies laser operation, that of light amplification by stimulated emission of radiation, which in turn gives rise to the acronym *LASER*.

In practice, a laser is generally a generator of radiation. The generator is constructed by housing a light-emitting medium within a cavity defined by mirrors that provide feedback of emitted radiation through the medium. With sustained excitation of the ionic or molecular species of the medium to give a large density of excited energy states, the spontaneous and attendant stimulated emission of radiation from these states by photons of identical wavelength (a lossless process), which is amplified by feedback due to photon reflection by the cavity mirrors, leads to the generation of a very large photon density within the cavity. With one cavity mirror being partially transmissive, say 0.1 to 1%, a fraction of the cavity energy is emitted as an intense beam. With suitable selection of a laser medium, cavity geometry, and peak wavelengths of mirror reflection, the beam is also essentially monochromatic and very nearly collimated.

Identity of the lasing molecular species or laser medium fixes the output wavelength of the laser. Laser media range from gases within a tubular cavity, organic dye molecules dissolved in a flowing inert liquid carrier and heat sink, to impurity-doped transparent crystalline rods (solid state lasers) and semiconducting diode junctions [Lengyel, 1971]. The different physical properties of these media in part determine the methods used to excite them into lasing states.

Gas-filled, or gas lasers are typically excited by dc or rf electric current. The current either ionizes and excites the lasing gas, e.g., argon, to give the electronically excited and lasing Ar+ion, or ionizes a gaseous species in a mixture also containing the lasing species, e.g., N_2, which by efficient energy transfer excites the lasing molecular vibrational states of the CO_2 molecule.

Dye lasers and so-called solid-state lasers are typically excited by intense light from either another laser or from a flash lamp. The excitation light wavelength range is selected to ensure efficient excitation at the absorption wavelength of the lasing species. Both excitation and output can be continuous, or the use of a pulsed flashlamp or pulsed exciting laser to pump a solid-state or dye laser gives pulsed output with high peak power and short pulse duration of 1 μs to 1 ms. Repeated excitation gives a train of pulses. Additionally, pulses of higher peak power and shorter duration of approximately 10 ns can be obtained from solid lasers by intracavity Q-switching [Lengyel, 1971]. In this method, the density of excited states is transiently greatly increased by impeding the path between the totally reflecting and partially transmitting mirror of the cavity interrupting the stimulated emission process. Upon rapid removal of the impeding device (a beam-interrupting or -deflecting device), stimulated emission of the very large population of excited lasing states leads to emission of an intense laser pulse. The process can give single pulses or can be repeated to give a pulse train with repetition frequencies typically ranging from 1 Hz to 1kHz.

Gallium aluminum (GaAlAs) lasers are, as are all semiconducting diode lasers, excited by electrical current that creates excited hole-electron pairs in the vicinity of the diode junction. Those carrier pairs are the lasing species that emit spontaneously and with photon stimulation. The beam emerges parallel to the function with the plane of the function forming the cavity and thin-layer surface mirrors providing reflection. Use of continuous or pulsed excitation current results in continuous or pulsed output.

29.9 Biomedical Laser Beam Delivery Systems

Beam delivery systems for biomedical lasers guide the laser beam from the output mirror to the site of action on tissue. Beam powers of up to 100 W are transmitted routinely. All biomedical lasers incorporate a coaxial aiming beam, typically from a HeNe laser (632.8 nm) to illuminate the site of incidence on tissue.

Usually, the systems incorporate two different beam-guiding methods, either (1) a flexible fused silica *(SiO_2)* optical fiber or light guide, generally available currently for laser beam wavelengths between ≅400 nm and ≅2.1 μm, where SiO_2 is essentially transparent and (2) an articulated arm having beam-guiding mirrors for wavelengths greater than circa 2.1 μm (e.g., CO_2 lasers), for the Er: YAG and for pulsed lasers having peak power outputs capable of causing damage to optical fiber surfaces due to ionization by the intense electric field (e.g., pulsed ruby). The arm comprises straight tubular sections articulated together with high-quality power-handling dielectric mirrors at each articulation junction to guide the beam through each of the sections. Fused silica optical fibers usually are limited to a length of 1 to 3 m and to wavelengths in the visible-to-low midrange IR (<2.1μm), because longer wavelengths of IR radiation are absorbed by water impurities (<2.9 μm) and by the SiO_2 lattice itself (wavelengths >5 μm), as described by Levi [1980].

Since the flexibility, small diameter, and small mechanical inertia of optical fibers allow their use in either flexible or rigid endoscopes and offer significantly less inertia to hand movement, fibers for use at longer IR wavelengths are desired by clinicians. Currently, researchers are evaluating optical fiber materials transparent to longer IR

wavelengths. Material systems showing promise are fused $A1_2O_3$ fibers in short lengths for use with near-3-micrometer radiation of the Er: YAG laser and *Ag halide* fibers in short lengths for use with the CO_2 laser emitting at 10.6 μm [Merberg, 1993]. A flexible, hollow Teflon waveguide 1.6 mm in diameter having a thin metal film overlain by a dielectric layer has been reported recently to transmit 10.6 μm CO_2 radiation with attenuation of 1.3 and 1.65 dB/m for straight and bent (5-mm radius, 90-degree bend) sections, respectively [Cannot et al., 1994].

Optical Fiber Transmission Characteristics

Guiding of the emitted laser beam along the optical fiber, typically of uniform circular cross-section, is due to total internal reflection of the radiation at the interface between the wall of the optical fiber core and the cladding material having refractive index n_1 less than that of the core n_2 [Levi, 1980]. Total internal reflection occurs for any angle of incidence θ of the propagating beam with the wall of the fiber core such that $\theta > \theta_c$, where

$$\sin \theta_c = \left(\frac{n_1}{n_2} \right), \tag{29.2}$$

or in terms of the complementary angle α_c,

$$\cos \alpha_c = \left(\frac{n_1}{n_2} \right). \tag{29.3}$$

For a focused input beam with apical angle α_m incident upon the flat face of the fiber as shown in Fig. 29.1, total internal reflection and beam guidance within the fiber core will occur [Levi, 1980] for

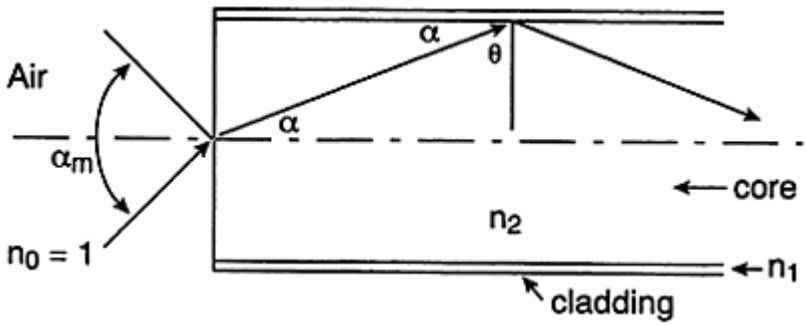


FIGURE 29.1 Critical reflection and propagation within an optical fiber.

$$NA = \sin\left(\alpha_m/2\right) \le \left[n_2^2 - n_1^2\right]^{0.5}, \tag{29.4}$$

where *NA* is the numerical aperture of the fiber.

This relationship ensures that the critical angle of incidence of the interface is not exceeded and that total internal reflection occurs [Levi, 1980]. Typical values of *NA* for fused SiO_2 fibers with polymer cladding are in the range of 0.36 to 0.40. The typical values of α_m=14 degrees used to insert the beam of the biomedical laser into the fiber is

$$\frac{n_1}{n_2} \leq \frac{1-\rho}{1+\rho}.$$

much smaller than those values (21 to 23 degrees) corresponding to typical *NA* values. The maximum value of the propagation angle oc typically used in biomedical laser systems is $\simeq$4.8 degrees.

Leakage of radiation at the core-cladding interface of the fused SiO_2 fiber is negligible, typically being $\simeq$0.3 dB/m at 400 nm and 0.01 dB/m at 1.064 μm. Bends along the fiber length always decrease the angle of the incidence at the core-cladding interface. Bends do not give appreciable losses for values of the bending radius sufficiently large that the angle of incidence θ of the propagating beam in the bent core does not becomes less than θ_c at the core-cladding interface [Levi, 1980]. The relationship given by Levi [1980] between the bending radius r_b, the fiber core radius r_o, the ratio (n_2/n_1) of fiber core to cladding refractive indices, and the propagation angle α in Fig. 29.1, which ensures that the beam does not escape is

$$\frac{n_1}{n_2} > \frac{1-\rho}{1+\rho} \cos\alpha, \tag{29.5}$$

where $\rho=(r_0/r_b)$. The inequality will hold for all $\alpha \leq \alpha_c$ provided that

$$\frac{n_1}{n_2} \leq \frac{1-\rho}{1+\rho}. \tag{29.6}$$

Thus, the critical bending radius r_{bc} is the value of r_b such that Eq. (29.6) is an equality. Use of Eq. (29.6) predicts that bends with radii ≥12, 18, and 30 mm, respectively, will not result in appreciable beam leakage from fibers having 400-, 600-, and 1000-micron diameter cores, which are typical in biomedical use. Thus, use of fibers in flexible endoscopes usually does not compromise beam guidance.

Because the integrity of the core-cladding interface is critical to beam guiding, the clad fiber is encased typically in a tough but flexible protective fluoropolymer buffer coat.

Mirrored, Articulated Arm Characteristics

Typically two or three relatively long tubular sections or arms of 50 to 80 cm length make up the portion of the articulated arm that extends from the laser output fixturing to the handpiece, endoscope, or operating microscope stage used to position the laser beam onto the tissue proper. Mirrors placed at the articulation of the arms and within the articulated handpiece, laparoscope, or operating microscope stage maintain the centration of the trajectory of the laser beam along the length of the delivery system. Dielectric multilayer mirrors [Levi, 1980] routinely are used in articulated devices. Their low high reflectivity≤99.9+% and power-handling capabilities ensure efficient power transmission

down the arm. Mirrors in articulated devices typically are held in kinetically adjustable mounts for rapid stable alignment to maintain beam concentration.

Optics for Beam Shaping on Tissues

Since the rate of heating on tissue, and hence rates of ablation and coagulation, depends directly on energy input per unit volume of tissue, selection of ablation and coagulation rates of various tissues is achieved through control of the energy density (J/cm^2 or $W \cdot s/cm^2$) of the laser beam. This parameter is readily achieved through use of optical elements such as discrete focusing lenses placed in the handpiece or rigid endoscope that control the spot size upon the tissue surface or by affixing a so-called contact tip to the end of an optical fiber. These are conical or spherical in shape with diameters ranging from 300 to 1200 μm and with very short focal lengths. The tip is placed in contact with the tissue and generates a submillimeter-sized focal spot in tissue very near the interface between the tip and tissue. One advantage of using the contact tip over a focused beam is that ablation proceeds with small lateral depth of attendant coagulation [Judy et al., 1993a]. This is because the energy of the tightly focused beam causes tissue thermolysis essentially at the tip surface and because the resulting tissue products strongly absorb the beam resulting in energy deposition and ablation essentially at the tip surface. This contrasts with the radiation penetrating deeply into tissue before thermolysis that occurs with a less tightly focused beam from a free lens or fiber. An additional advantage with the use of contact tips in the perception of the surgeon is that the kinesthetics of moving a contact tip along a resisting tissue surface more closely mimics the “touch” encountered in moving a scalpel across the tissue surface.

Recently a class of optical fiber tips has been developed that laterally directs the beam energy from a silica fiber [Judy et al., 1993b]. These tips, either a gold reflective micromirror or an angled refractive prism, offer a lateral angle of deviation ranging from 35 to 105 degrees from the optical fiber axis (undeviated beam direction). The beam reflected from a plane micromirror is unfocused and circular in cross-section, whereas the beam from a concave mirror and refractive devices is typically elliptical in shape, fused with distal diverging rays. Fibers with these terminations are currently finding rapidly expanding, large-scale application in coagulation (with 1.064-μm Nd: YAG laser radiation) of excess tissue lining the urethra in treatment of benign prostatic hypertrophy [Costello et al., 1992]. The capability for lateral beam direction may offer additional utility of these terminated fibers in other clinical specialties.

Features of Routinely Used Biomedical Lasers

Currently four lasers are in routine, large-scale clinical biomedical use to ablate, dissect, and to coagulate soft tissue. Two, the carbon dioxide (CO_2) and argon ion (Ar-ion) lasers, are gas-filled lasers. The other two employ solid-state lasing media. One is the Neodymium-yttrium-aluminum-garnet (Nd: YAG) laser, commonly referred to as a solid-state laser, and the other is the gallium-aluminum arsenide (GaAlAs) semiconductor diode laser. Salient features of the operating characteristics and biomedical applications of those lasers are listed in Tables 29.2 to 29.5. The operational descriptions are typical of

the lasers currently available commercially and do not represent the product of any single manufacturer.

TABLE 29.2 Operating Characteristics of Principal Biomedical Lasers

Characteristics	Ar Ion Laser	CO_2 Laser
Cavity medium	Argon gas, 133 Pa	10% CO_2 10% Ne, 80% He; 1330 Pa
Lasing species	Ar+ ion	CO_2 molecule
Excitation	Electric discharge, continuous	Electric discharge, continuous, pulsed
Electric input	208 V_{AC}, 60 A	110V_{AC}, 15 A
Wall plug efficiency	Φ0.06%	Φ10%
Characteristics	**Nd: YAG Laser**	**GaAlAs Diode Laser**
Cavity medium	Nd-doped YAG	n-p junction, GaAlAs diode
Lasing species	Nd3t in YAG lattice	Hole-electron pairs at diode junction
Excitation	Flashlamp, continuous, pulsed	Electric current, continuous pulsed
Electric input	208/240 V_{AC}, 30 A continuous 110V_{AC}, 10 A pulsed	110V_{AC}, 15 A
Wall plug efficiency	Φ1%	Φ23%

TABLE 29.3 Output Beam Characteristics of Ar-Ion and CO_2 Biomedical Lasers

Output Characteristics	Argon Laser	CO, Laser
Output power	2–8 W, continuous	1–100 W, continuous
Wavelength(s)	Multiple lines (454.6–528.7 nm), 488, 514.5 dominant	10.6μm
Electromagnetic wave propagation mode	TEM_{∞}	TEM_{∞}
Beam guidance, shaping	Fused silica optical fiber with contact tip or flat-ended for beam emission, lensed handpiece. Slit lamp with ocular lens	Flexible articulated arm with mirrors; lensed handpiece or mirrored microscope platen

TABLE 29.4 Output Beam Characteristics of Nd: YAG and GaAlAs Diode Biomedical Lasers

Output Characteristics	Nd: YAG Lasers	GaAlAs Diode Laser
Output power	1–100 W continuous at 1.064 millimicron 1–36 W continuous at 532 nm (frequency doubled with KTP)	1–25 W continuous
Wavelength(s)	1.064 μm/532 nm	810 nm
Electromagnetic wave propagation modes	Mixed modes	Mixed modes
Beam guidance and shaping	Fused SiO_2 optical fiber with contact tip directing mirrored or refracture tip	Fused SiO_2 optical fiber with contact tip or laterally directing mirrored or refracture tip

TABLE 29.5 Clinical Uses of Principal Biomedical Lasers

Ar-ion laser: Pigmented (vascular) soft-tissue ablation in gynecology; general and oral surgery; otolaryngology; vascular lesion coagulation in dermatology; retinal coagulation in ophthalmology	*CO_2 laser:* Soft-tissue ablation—dissection and bulk tissue removal in dermatology; gynecology; general, oral, plastic, and neurosurgery; otolaryngology; podiatry; urology
Nd: YAG laser: Soft-tissue, particularly pigmented vascular tissue, ablation—dissection and bulk tissue removal—in dermatology; gastroenterology; gynecology; general, arthroscopic, neuro-, plastic, and thoracic surgery; urology; posterior capsulotomy (ophthalmology) with pulsed 1.064 millimicron and ocular lens	*GaAlAs diode laser:* Pigmented (vascular) soft-tissue ablation—dissection and bulk removal in gynecology; gastroenterology, general surgery, and urology; FDA approval for otolaryngology and thoracic surgery pending

Other Biomedical Lasers

Some important biomedical lasers have smaller-scale use or currently are being researched for biomedical application. The following four lasers have more limited scales of surgical use:

The Ho: YAG (Holmium: YAG) laser, emitting pulses of 2.1-μm wavelength and up to 4 J in energy, used in soft tissue ablation in arthroscopic (joint) surgery (FDA approved).

The Q-switched Ruby *($Cr:Al_20_3$)* laser, emitting pulses of 694-nm wavelength and up to 2 J in energy is used in dermatology to disperse black, blue, and green tattoo pigments and melanin in pigmented lesions (not melanoma) for subsequent removal by phagocytosis by macrophages (FDA approved).

The flashlamp-pumped, pulsed-dye laser emitting 1-to 2-J pulses at either 577-or 585-nm wavelength (near the 537–577 absorption region of blood) is used for treatment of

cutaneous vascular lesions and melanin pigmented lesions except melanoma. Use of pulsed radiation helps to localize the thermal damage to within the lesions to obtain low damage of adjacent tissue.

The following lasers are being investigated for clinical uses.

The Er: YAG laser, emitting at 2.94 μm near the major water absorption peak (OH stretch), is currently being investigated for ablation of tooth enamel and dentin [Li et al., 1992].

Dye lasers emitting at 630 to 690 nm are being investigated for application as light sources for exciting dihematoporphyrin ether or benzoporphyrin derivatives in investigation of the efficacy of these photosensitives in the treatment of esophageal, bronchial, and bladder carcinomas for the FDA approved process.

Defining Terms

Biomedical Laser Radiation Ranges

Infrared (IR) radiation: The portion of the electromagnetic spectrum within the wavelength range 760 nm to 1 mm, with the regions 760 nm to 1.400 μm and 1.400 to 10.00 μm, respectively, called the near- and mid-IR regions.
Ultraviolet (UV) radiation: The portion of the electromagnetic spectrum within the wavelength range 100 to 400 nm.
Visible (VIS) radiation: The portion of the electromagnetic spectrum within the wavelength range 400 to 760 nm.

Laser Medium Nomenclature

Argon fluoride (ArF): Argon fluoride eximer laser (an eximer is a diatomic molecule that can exist only in an excited state).
Ar ion: Argon ion.
CO_2: Carbon dioxide.
Cr:Al2O3: Ruby laser.
Er: YAG: Erbium yttrium aluminum garnet.
GaAlAs: Gallium aluminum laser.
HeNe: Helium neon laser.
Ho: YAG: Holmium yttrium aluminum garnet.
Nd: YAG: Neodymium yttrium aluminum garnet.

Optical Fiber Nomenclature

Ag halide: Silver halide, halide ion, typically bromine (Br) and chlorine (Cl).
Fused silica: Fused SiO_2.

References

Bass LS, Moazami N, Pocsidio J, et al. 1992. Change in type I collagen following laser welding. *Lasers Surg Med* 12(5):500.

Costello AJ, Johnson DE, Bolton DM. 1992. Nd: YAG laser ablation of the prostate as a treatment for benign prostate hypertrophy. *Lasers Surg Med* 12(2):121.

Fitzpatrick RE. 1993. Comparison of the Q-switched ruby, Nd: YAG, and alexandrite lasers in tattoo removal. *Lasers Surg Med* Suppl 6(266):52.

Gannot I, Dror J, Calderon S, et al. 1994. Flexible waveguides for IR laser radiation and surgery applications. *Lasers Surg Med* 14(2):184.

Garrison BJ, Srinivasan R. 1985. Laser ablation of organic polymers: microscopic models for photochemical and thermal processes. *J Appl Physiol* 58(9):2909.

Grossweiner LI. 1989. Photophysics. In KC Smith (ed), *The Science of Photobiology*, p 1–47. New York, Plenum.

Hale GM, Querry MR. 1973. Optical constants of water in the 200 nm to 200 μm wavelength region. *Appl Optics* 12(12):555.

Hibst R, Bende T, Schroder D. 1992. Wet corneal ablation by Er: YAG laser radiation. *Lasers Surg Med* Suppl 4(236):56.

Hill RA, Le MT, Yashiro H, et al. 1993. Ab-interno erbium (Er: YAG) laser schlerostomy with iridotomy in dutch cross rabbits. *Lasers Surg Med* 13(5):559.

Judy MM, Matthews JL, Aronoff BL, et al. 1993a. Soft tissue studies with 805 nm diode laser radiation: Thermal effects with contact tips and comparison with effects of 1064 nm. Nd: YAG laser radiation. *Lasers Surg Med* 13(5):528.

Judy MM, Matthews JL, Gardetto WW, et al. 1993b. Side firing laser-fiber technology for minimally invasive transurethral treatment of benign prostate hyperplasia. *Proc Soc Photo-Optical Instr Eng* (SPIE) 1982:86.

Judy MM, Matthews JL, Goodson JR, et al. 1992. Thermal effects in tissues from simultaneous coaxial CO_2 and Nd: YAG laser beams. *Lasers Surg Med* 12(2):222.

Katoh N, Peyman GA. 1988. Effects of laser wavelengths on experimental retinal detachments and retinal vessels. *Jpn J Ophthalmol* 32(2): 196.

Keye WR, Matson GA, Dixon J. 1983. The use of the argon laser in treatment of experimental endometriosis. *Fertil Steril* 39(1):26.

Lakowicz JR. 1983. *Principles of Fluorescence Spectroscopy.* New York, Plenum.

Lengyel BA. 1971. *Lasers.* New York, John Wiley.

Levi L. 1980. *Applied Optics*, vol 2. New York, John Wiley.

Li ZZ, Code JE, Van de Merve WP. 1992. Er: YAG laser ablation of enamel and dentin of human teeth: determination of ablation rates at various fluences and pulse repetition rates. *Lasers Surg Med* 12(6):625.

Maes KE, Sherk HH. 1994. Bone and meniscal ablation using the erbium YAG laser. *Lasers Surg Med* Suppl 6(166):31.

Mahadevan A, Mitchel MF, Silva E, et al. 1993. Study of the fluorescence properties of normal and neoplastic human cervical tissue. *Lasers Surg Med* 13(6):647.

McGillis ST, Bailin PL, Fitzpatrick RE, et al. 1994. Successful treatments of blue, green, brown and reddishbrown tattoos with the Q-switched alexandrite laser. *Lasers Surg Med Suppl* 6(270):52.

Meijering LJT, VanGermert MJC, Gijsbers GHM, et al. 1993. Limits of radial time constants to approximate thermal response of tissue. *Lasers Surg Med* 13(6):685.

Merberg GN. 1993. Current status of infrared fiberoptics for medical laser power delivery. *Lasers Surg Med* 13(5):572.

Miller ID, Veitch AR. 1993. Optical modeling of light distributions in skin tissue following laser irradiation. *Lasers Surg Med* 13(5):565.

Mordon S, Beacco C, Rotteleur G, et al. 1993. Relation between skin surface temperature and minimal blanching during argon, Nd: YAG 532, and cw dye 585 laser therapy of port-wine stains. *Lasers Surg Med* 13(1):124.

Motamedi M, Torres JH, Cammack T, et al. 1993. Thermodynamics of cw laser interaction with prostatic tissue: Effects of simultaneous cooling on lesion size. *Lasers Surg Med Suppl* 5(314):64.

Oz MC, Chuck RS, Johnson JP, et al. 1989. Indocyanine green dye-enhanced welding with a diode laser. *Surg Forum* 40(4):316.

Rastegar S, Jacques SC, Motamedi M, et al. 1992. Theoretical analysis of high-power diode laser (810 nm) and Nd: YAG laser (1064 nm) for coagulation of tissue: Predictions for prostate coagulation. *Proc Soc Photo-Optical Instr Eng* (SPIE) 1646:150.

Schroder T, Brackett K, Joffe S. 1987. An experimental study of effects of electrocautery and various lasers on gastrointestinal tissue. *Surgery* 101(6):691.

Spikes JD. 1989. Photosensitization. In KG Smith (ed), *The Science of Photobiology,* 2d ed, pp 79–110. New York, Plenum.

Van de Hulst HC. 1957. *Light Scattering by Small Particles.* New York, John Wiley.

Van Saarloos PP, Constable IJ. 1993. Improved eximer laser photorefractive keratectomy system. *Lasers Surg Med* 13(2):189.

White A, Handler P, Smith EL. 1968. *Principles of Biochemistry,* 4th ed. New York, McGraw-Hill.

Further Information

Current research on the optical, thermal, and photochemical interactions of radiation and their effect on biologic tissues, are published routinely in the journals: *Laser in Medicine and Surgery*, *Lasers in the Life Sciences,* and *Photochemistry Photobiology* and to a lesser extent in *Applied Optics and Optical Engineering*.

Clinical evaluations of biomedical laser applications appear in *Lasers and Medicine and Surgery* and in journals devoted to clinical specialties such as *Journal of General Surgery, Journal of Urology, Journal of Gastroenterological Surgery*.

The annual symposium proceedings of the biomedical section of the Society of Photo-Optical Instrumentation Engineers (SPIE) contain descriptions of new and current research on application of lasers and optics in biomedicine.

The book *Lasers* (a second edition by Bela A. Lengyel), although published in 1971, remains a valuable resource on the fundamental physics of lasers—gas, dye solid-state, and semiconducting diode. A more recent book, *The Laser Guidebook*, by Jeffrey Hecht, published in 1992, emphasizes the technical characteristics of the gas, diode, solid-state, and semiconducting diode lasers.

The *Journal of Applied Physics, Physical Review Letters,* and *Applied Physics Letters* carry descriptions of the newest advances and experimental phenomena in lasers and optics.

The book *Safety with Lasers and Other Optical Sources* by David Sliney and Myron Wolbarsht, published in 1980, remains a very valuable resource on matters of safety in laser use.

Laser safety standards for the United States are given for all laser uses and types in the American National Standard (ANSI) Z136.1–1993, Safe Use of Lasers.

30
Noninvasive Optical Monitoring

Ross Flewelling
Nellcor Incorporation

Optical measures of physiologic status are attractive because they can provide a simple, noninvasive, yet real-time assessment of medical condition. Noninvasive optical monitoring is taken here to mean the use of visible or near-infrared light to directly assess the internal physiologic status of a person without the need of extracting a blood of tissue sample or using a catheter. Liquid water strongly absorbs ultraviolet and infrared radiation, and thus these spectral regions are useful only for analyzing thin surface layers or respiratory gases, neither of which will be the subject of this review. Instead, it is the visible and near-infrared portions of the electromagnetic spectrum that provide a unique "optical window" into the human body, opening new vistas for noninvasive monitoring technologies.

Various molecules in the human body possess distinctive spectral absorption characteristics in the visible or near-infrared spectral regions and therefore make optical monitoring possible. The most strongly absorbing molecules at physiologic concentrations are the hemoglobins, myoglobins, cytochromes, melanins, carotenes, and bilirubin (see Fig. 30.1 for some examples). Perhaps less appreciated are the less distinctive and weakly absorbing yet ubiquitous materials possessing spectral characteristics in the near-infrared: water, fat, proteins, and sugars. Simple optical methods are now available to quantitatively and noninvasively measure some of these compounds directly in intact tissue. The most successful methods to date have used hemoglobins to assess the oxygen content of blood, cytochromes to assess the respiratory status of cells, and possibly near-infrared to assess endogenous concentrations of metabolytes, including glucose.

30.1 Oximetry and Pulse Oximetry

Failure to provide adequate oxygen to tissues—*hypoxia*—can in a matter of minutes result in reduced work capacity of muscles, depressed mental activity, and ultimately cell death. It is therefore of considerable interest to reliably and accurately determine the amount of oxygen in blood or tissues. *Oximetry* is the determination of the oxygen content of blood of tissues, normally by optical means. In the clinical laboratory the oxygen content of whole blood can be determined by a bench-top cooximeter or blood-

0-8493-1813-0/03/$0.00+$1.50

gas analyzer. But the need for timely clinical information and the desire to minimize the inconvenience and cost of extracting a blood sample and later analyze it in the lab has led to the search for alternative

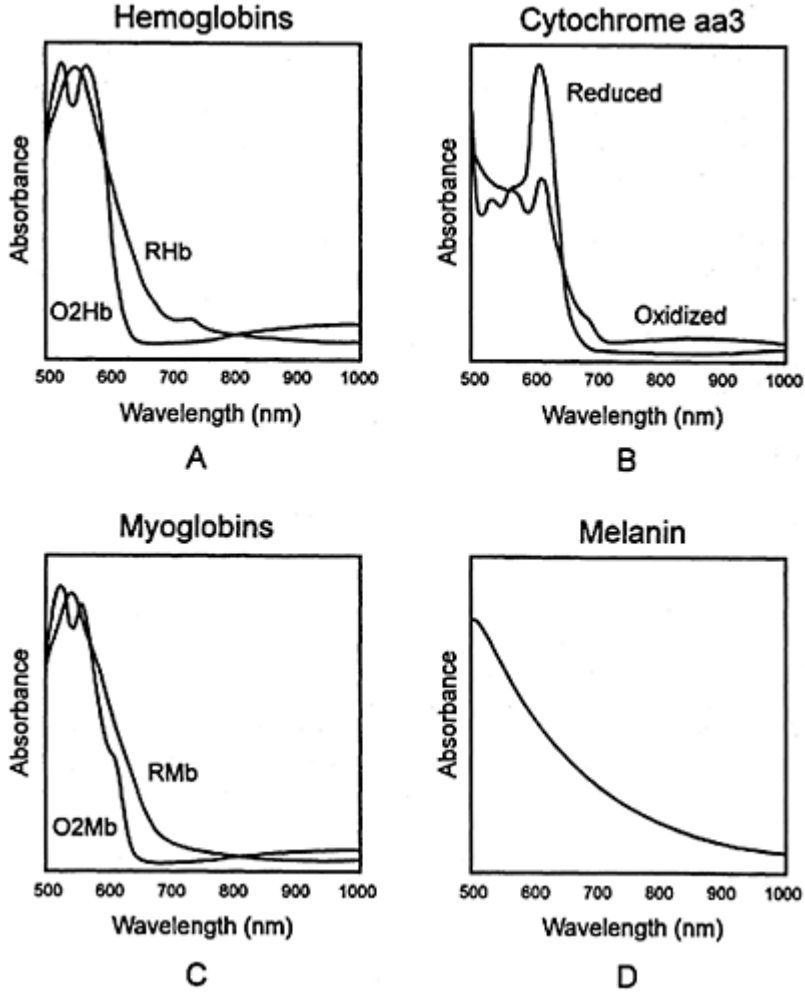


FIGURE 30.1 Absorption spectra of some endogenous biologic materials *(a)* hemoglobins, *(b)* cytochrome aa3, *(c)* myoglobins, and *(d)* melanin.

noninvasive optical methods. Since the 1930s, attempts have been made to use multiple wavelengths of light to arrive at a complete spectral characterization of a tissue. These

approaches, although somewhat successful, have remained of limited utility owing to the awkward instrumentation and unreliable results.

It was not until the invention of *pulse oximetry* in the 1970s and its commercial development and application in the 1980s that noninvasive oximetry became practical. Pulse oximetry is an extremely easy-to-use, noninvasive, and accurate measurement of real-time arterial oxygen saturation. Pulse oximetry is now used routinely in clinical practice, has become a standard of care in all U.S. operating rooms, and is increasingly used wherever critical patients are found. The explosive growth of this new technology and its considerable utility led John Severinghaus and Poul Astrup [1986] in an excellent historical review to conclude that pulse oximetry was "arguably the most significant technological advance ever made in monitoring the well-being and safety of patients during anesthesia, recovery and critical care."

Background

The partial pressure of oxygen (pO_2) in tissues need only be about 3 mmHg to support basic metabolic demands. This tissue level, however, requires capillary pO_2 to be near 40 mmHg, with a corresponding

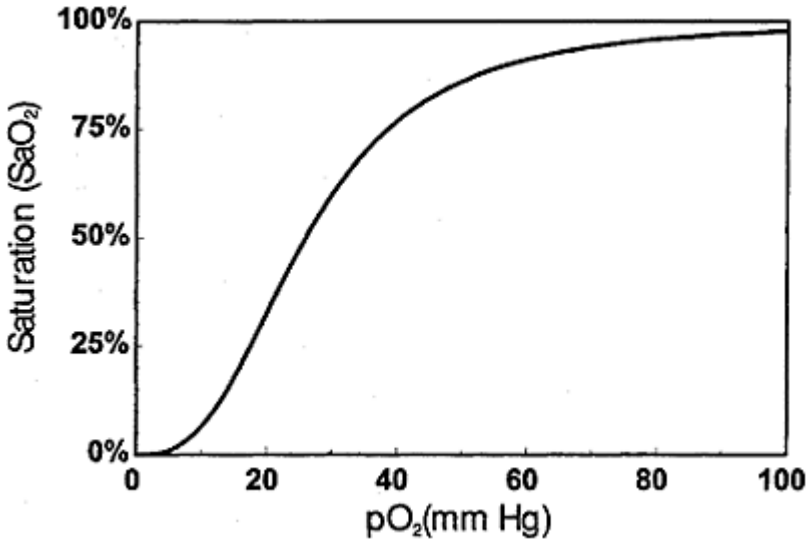


FIGURE 30.2 Hemoglobin oxygen dissociation curve showing the sigmoidal relationship between the partial pressure of oxygen and the oxygen saturation of blood. The curve is given approximately by

$$\%SaO_2 = 100\%/[1+P_{50}/\ pO_2^{n}], \text{ with}$$
$$n=2.8 \text{ and } P_{50}=26 \text{ mmHg}.$$

arterial pO_2 of about 95 mmHg. Most of the oxygen carried by blood is stored in red blood cells reversibly bound to hemoglobin molecules. Oxygen saturation (SaO_2) is defined as the percentage of hemoglobin-bound oxygen compared to the total amount of hemoglobin available for reversible oxygen binding. The relationship between the oxygen partial pressure in blood and the oxygen saturation of blood is given by the hemoglobin oxygen dissociation curve as shown in Fig. 30.2. The higher the pO_2 in blood, the higher the SaO_2. But due to the highly cooperative binding of four oxygen molecules to each hemoglobin molecule, the oxygen binding curve is sigmoidal, and consequently the SaO_2 value is particularly sensitive to dangerously low pO_2 levels. With a normal arterial blood pO_2 above 90 mmHg, the oxygen saturation should be at least 95%, and a pulse oximeter can readily verify a safe oxygen level. If oxygen content falls, say to a pO_2 below 40 mmHg, metabolic needs may not be met, and the corresponding oxygen saturation will drop below 80%. Pulse oximetry therefore provides a direct measure of oxygen sufficiency and will alert the clinician to any danger of imminent hypoxia in a patient.

Although endogenous molecular oxygen is not optically observable, hemoglobin serves as an oxygen-sensitive "dye" such that when oxygen reversibly binds to the iron atom in the large heme prosthetic group, the electron distribution of the heme is shifted, producing a significant color change. The optical absorption of hemoglobin in its oxygenated and deoxygenated states is shown in Fig. 30.1. Fully oxygen-ated blood absorbs strongly in the blue and appears bright red; deoxygenated blood absorbs through the visible region and is very dark (appearing blue when observed through tissue due to light-scattering effects). Thus the optical absorption spectra of oxyhemaglobin (O_2Hb) and "reduced" deoxyhemoglobin (RHb) differ substantially, and this difference provides the basis for spectroscopic determinations of the proportion of the two hemoglobin states. In addition to these two normal functional hemoglobins, there are also *dysfunctional hemoglobins*—carboxyhemoglobin, methemoglobin, and sulhemoglobin—which are spectroscopically distinct but do not bind oxygen reversibly. Oxygen saturation is therefore defined in Eq. (30.1) only in terms of the *functional saturation* with respect to O_2Hb and RHb:

$$S_aO_2 = \frac{O_2Hb}{RHb+O_2Hb} \times 100\% . \qquad (30.1)$$

Cooximeters are bench-top analyzers that accept whole-blood samples and utilize four or more wavelengths of monochromatic light, typically between 500 and 650 nm, to spectroscopically determine the various individual hemoglobins in the sample. If a blood sample can be provided, this spectroscopic method is accurate and reliable. Attempts to make an equivalent quantitative analysis noninvasively through intact tissue have been fraught with difficulty. The problem has been to contend with the wide variation in scattering and nonspecific absorption properties of very complex heterogeneous tissue. One of the more successful approaches, marketed by Hewlett-Packard, used eight optical wavelengths transmitted through the pinna of the ear. In this approach a "bloodless"

measurement is first obtained by squeezing as much blood as possible from an area of tissue; the arterial blood is then allowed to flow back, and the oxygen saturation is determined by analyzing the change in the spectral absorbance characteristics of the tissue. While this method works fairly well, it is cumbersome, operator dependent, and does not always work well on poorly perfused or highly pigmented subjects.

In the early 1970s, Takuo Aoyagi recognized that most of the interfering nonspecific tissue effects could be eliminated by utilizing only the change in the signal during an arterial pulse. Although an early prototype was built in Japan, it was not until the refinements in implementation and application by Biox (now Ohmeda) and Nellcor Incorporated in the 1980s that the technology became widely adopted as a safety monitor for critical care use.

Theory

Pulse oximetry is based on the fractional change in light transmission during an arterial pulse at two different wavelengths. In this method the fractional change in the signal is due only to the arterial blood itself, and therefore the complicated nonpulsatile and highly variable optical characteristics of tissue are eliminated. In a typical configuration, light at two different wavelengths illuminating one side of a finger will be detected on the other side, after having traversed the intervening vascular tissues (Fig. 30.3). The transmission of light at each wavelength is a function of the thickness, color, and structure of the skin, tissue, bone, blood, and other material through which the light passes. The absorbance of light by a sample is defined as the negative logarithm of the ratio of the light intensity in the presence of the sample *(I)* to that without *(I_o): $A=-\log(I/I_o)$*. According to the *Beer-Lambert law*, the absorbance of a sample at a given wavelength with a molar absorptivity (Φ) is directly proportional to both the concentration *(c)* and pathlength (l) of the absorbing material: $A=\Phi cl$. (In actuality, biologic tissue is highly scattering, and the Beer-Lambert law is only approximately correct; see the references for further elaboration). Visible or near-infrared light passing through about one centimeter of tissue (e.g., a finger) will be attenuated by about one or two orders of magnitude for a typical emitter-detector geometry, corresponding to an effective optical density (OD) of 1–2 OD (the detected light intensity is decreased by one order of

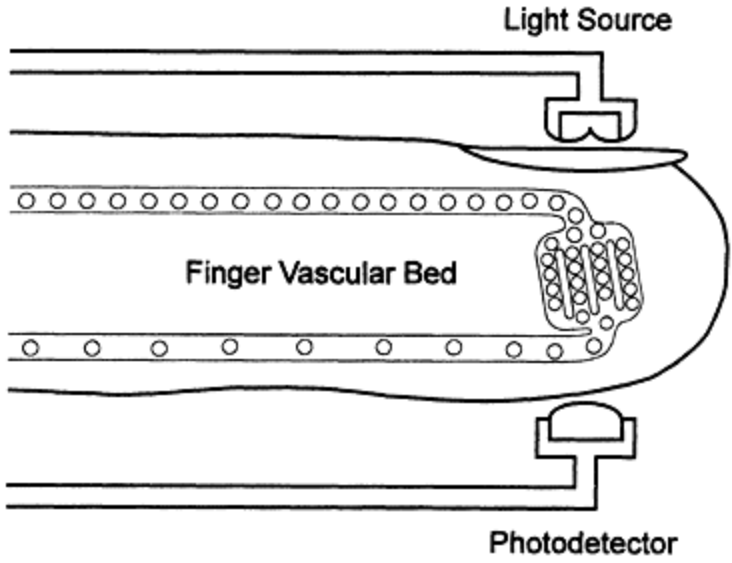


FIGURE 30.3 Typical pulse oximeter sensing configuration on a finger. Light at two different wavelengths is emitted by the source, diffusely scattered through the finger, and detected on the opposite side by a photodetector.

magnitude for each OD unit). Although hemoglobin in the blood is the single strongest absorbing molecule, most of the total attenuation is due to the scattering of light away from the detector by the highly heterogeneous tissue. Since human tissue contains about 7% blood, and since blood contains typically about 14 g/dL hemoglobin, the effective hemoglobin concentration in tissue is about 1 g/dL (~150 uM). At the wavelengths used for pulse oximetry (650–950 nm), the oxy- and deoxyhemoglobin molar absorptivities fall in the range of 100 to 1000 $M^{-1}\,cm^{-1}$, and consequently hemoglobin accounts for less than 0.2 OD of the total observed optical density. Of this amount, perhaps only 10% is pulsatile, and consequently pulse signals of only a few percent are ultimately measured, at times even one tenth of this.

A mathematical model for pulse oximetry begins by considering light at two wavelengths, λ_1 and λ_2, passing through tissue and being detected at a distant location as in Fig. 30.3. At each wavelength the total light attenuation is described by four different component absorbances: oxyhemoglobin in the blood (concentration c_0, molar absorptivity $d = cT_{,o}$ and effective pathlength l_o), "reduced" deoxyhemoglobin in the blood (concentration c_r, molar absorptivity $d = cT_{,r}$, and effective pathl ength l_r), specific variable absorbances that are not from the arterial blood (concentration c_x, molar

absorptivity $d = cT$ ›$_x$, and effective pathlength l_x), and all other nonspecific sources of optical attenuation, combined as A_y, which can include light scattering, geometric factors, and characteristics of the emitter and detector elements. The total absorbance at the two wavelengths can then be written:

$$\begin{cases} A_{\lambda_1} = \epsilon_{o_1} c_o l_o + \epsilon_{r_1} c_r l_r + \epsilon_{x_1} c_x l_x + A_{y_1} \\ A_{\lambda_2} = \epsilon_{o_2} c_o l_o + \epsilon_{r_2} c_r l_r + \epsilon_{x_2} c_x l_x + A_{y_2} \end{cases} . \tag{30.2}$$

The blood volume change due to the arterial pulse results in a modulation of the measured absorbances. By taking the time rate-of-change of the absorbances, the two last terms in each equation are effectively zero, since the concentration and effective pathlength of absorbing material outside the arterial blood do not change during a pulse [$d(c_x l_x)/dt$=0], and all the nonspecific effects on light attenuation are also effectively invariant on the time scale of a cardiac cycle (dA_y/dt=0). Since the extinction coefficients are constant, and the blood concentrations are constant on the time scale of a pulse, the time-dependent changes in the absorbances at the two wavelengths can be assigned entirely to the change in the blood pathlength (dl_o/dt and dl_r/dt). With the additional assumption that these two blood pathlength changes are equivalent (or more generally, their ratio is a constant), the ratio R of the time rate-of-change of the absorbance at wavelength *1* to that at wavelength *2* reduces to the following:

$$R = \frac{dA_{\lambda_1}/dt}{dA_{\lambda_2}/dt} = \frac{-d\log\left(I_1/I_o\right)/dt}{-d\log\left(I_2/I_o\right)/dt} = \frac{\left(\Delta I_1/I_1\right)}{\left(\Delta I_2/I_2\right)} = \frac{\epsilon_{o_1} c_o + \epsilon_{r_1} c_r}{\epsilon_{o_2} c_o + \epsilon_{r_2} c_r} . \tag{30.3}$$

Observing that functional oxygen saturation is given by $S=c_o/(c_o+c_r)$, and that $(1-S)=c_r/(c_o+c_r)$, the oxygen saturation can then be written in terms of the ratio R as follows:

$$S = \frac{\epsilon_{r1} - \epsilon_{r2} R}{\left(\epsilon_{r1} - \epsilon_{o1}\right) - \left(\epsilon_{r2} - \epsilon_{o2}\right) R} . \tag{30.4}$$

Equation (30.4) provides the desired relationship between the experimentally determined ratio R and the clinically desired oxygen saturation S. In actual use, commonly available LEDs are used as the light sources, typically a red LED near 660 nm and a near-infrared LED selected in the range 890 to 950 nm. Such LEDs are not monochromatic light sources, typically with bandwidths between 20 and 50 nm, and therefore standard molar absorptivities for hemoglobin cannot be used directly in Eq. (30.4). Further, the simple model presented above is only approximately true; for example, the two wavelengths

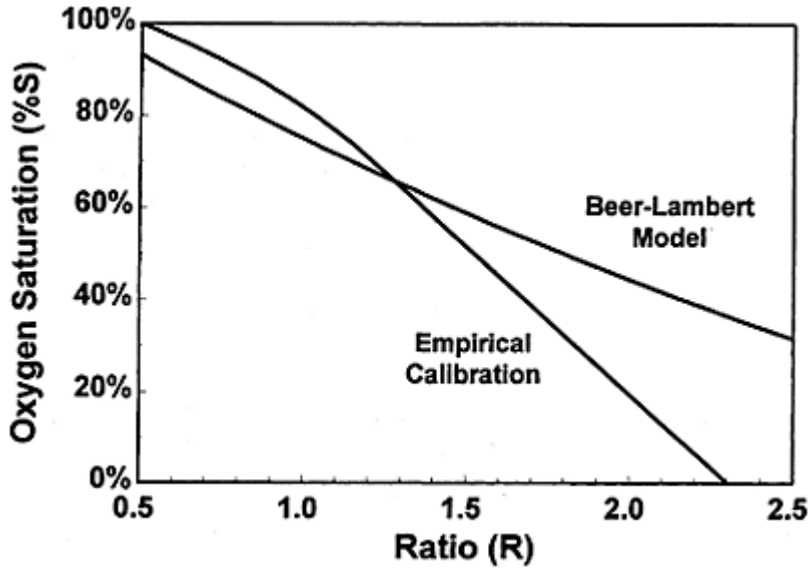


FIGURE 30.4 Relationship between the measured ratio of fractional changes in light intensity at two wavelengths, *R*, and the oxygen saturation S.Beer-Lambert model is from Eq. (30.4) with Φ_{o1}=100, Φ_{o2}=300, Φ_{r1}=800, and Φ_{r2}= 200. Empirical calibration is based on % S=100%×*(a—bR)/(c—dR)* with *a*=1000, *b*=550, *c*=900, and *d*=350, with a linear extrapolation below 70%.

do not necessarily have the exact same pathlength changes, and second-order scattering effects have been ignored. Consequently the relationship between *S* and *R* is instead determined empirically by fitting the clinical data to a generalized function of the form *S=(a—bR)/(c—dR)*. The final empirical calibration will ultimately depend on the details of an individual sensor design, but these variations can be determined for each sensor and included in unique calibration parameters. A typical empirical calibration for *R* versus *S* is shown in Fig. 30.4, together with the curve that standard molar absorptivities would predict.

In this way the measurement of the ratio of the fractional change in signal intensity of the two LEDs is used along with the empirically determined calibration equation to obtain a beat-by-beat measurement of the arterial oxygen saturation in a perfused tissue—continuously, noninvasively, and to an accuracy of a few percent.

Applications and Future Directions

Pulse oximetry is now routinely used in nearly all operating rooms and critical care areas in the United States and increasingly throughout the world. It has become so pervasive and useful that it is now being called the “fifth” vital sign (for an excellent review of practical aspects and clinical applications of the technology see Kelleher [1989]).

The principal advantages of pulse oximetry are that it provides continuous, accurate, and reliable monitoring of arterial oxygen saturation on nearly all patients, utilizing a variety of convenient sensors, reusable as well as disposable. Single-patient-use adhesive sensors can easily be applied to fingers for adults and children and to arms for legs or neonates. Surface reflectance sensors have also been developed based on the same principles and offer a wider choice for sensor location, though they tend to be less accurate and prone to more types of interference.

Limitations of pulse oximetry include sensitivity to high levels of optical or electrical interference, errors due to high concentrations of dysfunctional hemoglobins (methemoglobin or carboxyhemoglobin) or interference from physiologic dyes (such as methylene blue). Other important factors, such as total hemoglobin content, fetal hemoglobin, or sickle cell trait, have little or no effect on the measurement except under extreme conditions. Performance can also be compromised by poor signal quality, as may occur for poorly perfused tissues with weak pulse amplitudes or by motion artifact.

Hardware and software advances continue to provide more sensitive signal detection and filtering capabilities, allowing pulse oximeters to work better on more ambulatory patients. Already some pulse oximeters incorporate EGG synchronization for improved signal processing. A pulse oximeter for use in labor and delivery is currently under active development by several research groups and companies. A likely implementation may include use of a reflectance surface sensor for the fetal head to monitor the adequacy of fetal oxygenation. This application is still in active development, and clinical utility remains to be demonstrated.

30.2 Nonpulsatile Spectroscopy

30.2.1 Background

Nonpulsatile optical Spectroscopy has been used for more than half a century for noninvasive medical assessment, such as in the use of multiwavelength tissue analysis for oximetry and skin reflectance measurement for bilirubin assessment in jaundiced neonates. These early applications have found some limited use, but with modest impact. Recent investigations into new nonpulsatile Spectroscopy methods for assessment of deep-tissue oxygenation (e.g., cerebral-oxygen monitoring), for evaluation of respiratory status at the cellular level, and for the detection of other critical analytes, such as glucose, may yet prove more fruitful. The former applications have led to spectroscopic studies of cytochromes in tissues, and the latter has led to considerable work into new approaches in near-infrared analysis of intact tissues.

Cytochrome Spectroscopy

Cytochromes are electron-transporting, heme-containing proteins found in the inner membranes of mitochondria and are required in the process of oxidative phosphorylation to convert metabolytes and oxygen into CO_2 and high-energy phosphates. In this metabolic process the cytochromes are reversibly oxidized and reduced, and consequently the oxidation-reduction states of cytochromes *c* and aa_3 in particular are direct measures of the respiratory condition of the cell. Changes in the absorption spectra of these molecules, particularly near 600 nm and 830 nm for cytochrome aa_3, accompany this shift. By monitoring these spectral changes, the cytochrome oxidation state in the tissues can be determined (see, for example, Jöbsis [1977] and Jöbsis et al. [1977]). As with all nonpulsatile approaches, the difficulty is to remove the dependence of the measurements on the various nonspecific absorbing materials and highly variable scattering effects of the tissue. To date, instruments designed to measure cytochrome spectral changes can successfully track relative changes in brain oxygenation, but absolute quantitation has not yet been demonstrated.

Near-Infrared Spectroscopy and Glucose Monitoring

Near-infrared (NIR), the spectral region between 780 nm and 3000 nm, is characterized by broad and overlapping spectral peaks produced by the overtones and combinations of infrared vibrational modes. Figure 30.5 shows typical NIR absorption spectra of fat, water, and starch. Exploitation of this spectral region for *in vivo* analysis has been hindered by the same complexities of nonpulsatile tissue Spectroscopy described above and is further confounded by the very broad and indistinct spectral features characteristic of the NIR. Despite these difficulties, NIR Spectroscopy has garnered considerable attention, since it may enable the analysis of common analytes.

Karl Norris and coworkers pioneered the practical application of NIR spectroscopy, using it to evaluate water, fat, and sugar content of agricultural products (see Osborne et al. [1993] and Burns and Cuirczak [1992]). The further development of sophisticated *multivariate analysis* techniques, together with new scattering models (e.g., Kubelka-Munk theory) and high-performance instrumentation, further extended the application of NIR methods. Over the past decade, many research groups and companies have touted the use of NIR techniques for medical monitoring, such as for determining the relative fat, protein, and water content of tissue, and more recently for noninvasive glucose measurement. The body composition analyses are useful but crude and are mainly limited to applications in nutrition and sports medicine. Noninvasive glucose monitoring, however, is of considerable interest.

More than 2 million diabetics in the United States lance their fingers three to six times a day to obtain a drop of blood for chemical glucose determination. The ability of these individuals to control their glucose levels, and the quality of their life generally, would dramatically improve if a simple, noninvasive method for determining blood glucose levels could be developed. Among the noninvasive optical methods proposed for this purpose are optical rotation, NIR analysis, and Raman spectroscopy. The first two have received the most attention. Optical rotation methods aim to exploit the small optical rotation of polarized light by glucose. To measure physiologic glucose levels in a 1-cm

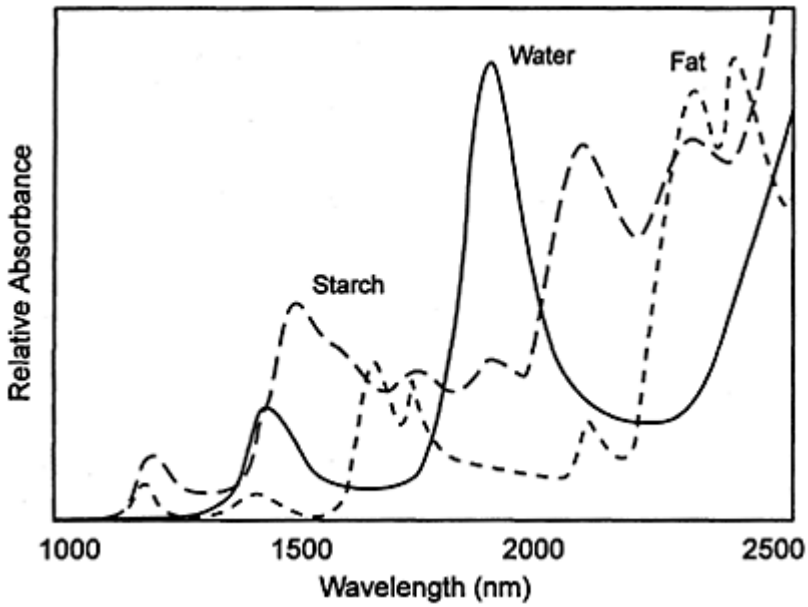


FIGURE 30.5 Typical near-infrared absorption spectra of several biologic materials.

thick sample to an accuracy of 25 mg/ dL would require instrumentation that can reliably detect an optical rotation of at least 1 millidegree. Finding an appropriate *in vivo* optical path for such measurements has proved most difficult, with most approaches looking to use either the aqueous humor or the anterior chamber of the eye [Coté et al., 1992; Rabinovitch et al., 1982]. Although several groups have developed laboratory analyzers that can measure such a small effect, so far *in vivo* measurement has not been demonstrated, due both to unwanted scattering and optical activity of biomaterials in the optical path and to the inherent difficulty in developing a practical instrument with the required sensitivity.

NIR methods for noninvasive glucose determination are particularly attractive, although the task is formidable. Glucose has spectral characteristics near 1500 nm and in the 2000 to 2500 nm band where many other compounds also absorb, and the magnitude of the glucose absorbance in biologic samples is typically two orders of magnitude lower than those of water, fat, or protein. The normal detection limit for NIR spectroscopy is on the order of one part in 10^3, whereas a change of 25 mg/dL in glucose concentration corresponds to an absorbance change of 10^{-4} to10^{-5}. In fact, the temperature dependence of the NIR absorption of water alone is at least an order of magnitude greater than the signal from glucose in solution. Indeed, some have suggested that the apparent glucose signature in complex NIR spectra may actually be the secondary effect of glucose on the water.

Sophisticated chemometric (particularly multivariate analysis) methods have been employed to try to extract the glucose signal out of the noise (for methods reviews, see

Martens and Næs [1989] and Haaland [1992]). Several groups have reported using multivariate techniques to quantitate glucose in whole-blood samples, with encouraging results [Haaland et al., 1992]. And despite all theoretical disputations to the contrary, some groups claim the successful application of these multivariate analysis methods to noninvasive *in vivo* glucose determination in patients [Robinson et al., 1992]. Yet even with the many groups working in this area, much of the work remains unpublished, and few if any of the reports have been independently validated.

Time-Resolved Spectroscopy

The fundamental problem in making quantitative optical measurements through intact tissue is dealing with the complex scattering phenomena. This scattering makes it difficult to determine the effective path length for the light, and therefore attempts to use the Beer-Lambert law, or even to determine a consistent empirical calibration, continue to be thwarted. Application of new techniques in time-resolved spectroscopy may be able to tackle this problem. Thinking of light as a packet of photons, if a single packet from a light source is sent through tissue, then a distant receiver will detect a photon distribution over time—the photons least scattered arriving first and the photons most scattered arriving later. In principle, the first photons arriving at the detector pass directly through the tissue. For these first photons the distance between the emitter and the detector is fixed and known, and the Beer-Lambert law should apply, permitting determination of an *absolute* concentration for an absorbing component. The difficulty in this is, first, that the measurement time scale must be on the order of the photon transit time (subnanosecond), and second, that the number of photons getting through without scattering will be extremely small, and therefore the detector must be exquisitely sensitive. Although these considerable technical problems have been overcome in the laboratory, their implementation in a practical instrument applied to a real subject remains to be demonstrated. This same approach is also being investigated for noninvasive optical imaging, since the unscattered photons should produce sharp images (see Chance et al., [1988], Chance [1991], and Yoo and Alfano [1989]).

30.3 Conclusions

The remarkable success of pulse oximetry has established noninvasive optical monitoring of vital physiologic functions as a modality of considerable value. Hardware and algorithm advances in pulse oximetry are beginning to broaden its use outside the traditional operating room and critical care areas. Other promising applications of noninvasive optical monitoring are emerging, such as for measuring deep tissue oxygen levels, determining cellular metabolic status, or for quantitative determination of other important physiologic parameters such as blood glucose. Although these latter applications are not yet practical, they may ultimately impact noninvasive clinical monitoring just as dramatically as pulse oximetry.

Defining Terms

Beer-Lambert law: Principle stating that the optical absorbance of a substance is proportional to both the concentration of the substance and the pathlength of the sample.
Cytochromes: Heme-containing proteins found in the membranes of mitochondria and required for oxidative phosphorylation, with characteristic optical absorbance spectra.
Dysfunctional hemoglobins: Those hemoglobin species that cannot reversibly bind oxygen (carboxyhemoglobin, methemoglobin, and sulfhemoglobin).
Functional saturation: The ratio of oxygenated hemoglobin to total nondysfunctional hemoglobins (oxyhemoglobin plus deoxyhemoglobin).
Hypoxia: Inadequate oxygen supply to tissues necessary to maintain metabolic activity.
Multivariate analysis: Empirical models developed to relate multiple spectral intensities from many calibration samples to known analyte concentrations, resulting in an optimal set of calibration parameters.
Oximetry: The determination of blood or tissue oxygen content, generally by optical means.
Pulse oximetry: The determination of functional oxygen saturation of pulsatile arterial blood by ratiometric measurement of tissue optical absorbance changes.

References

Burns DA, Ciurczak EW (eds). 1992. *Handbook of Near-Infrared Analysis.* New York, Marcel Dekker.
Chance B. 1991. Optical method. *Annu Rev Biophys Biophys Chem* 20:1.
Chance B, Leigh JS, Miyake H, et al. 1988. Comparison of time-resolved and -unresolved measurements of deoxyhemoglobin in brain. *Proc Natl Acad Sci USA* 85(14):4971.
Coté GL, Fox MD, Northrop RB. 1992. Noninvasive optical polarimetric glucose sensing using a true phase measurement technique. *IEEE Trans Biomed Eng* 39(7):752.
Haaland DM. 1992. Multivariate calibration methods applied to the quantitative analysis of infrared spectra. In PC Jurs (ed), *Computer-Enhanced Analytical Spectroscopy,* vol 3, pp 1–30. New York, Plenum.
Haaland DM, Robinson MR, Koepp GW, et al. 1992. Reagentless near-infrared determination of glucose in whole blood using multivariate calibration. *Appl Spectros* 46(10): 1575.
Jöbsis FF. 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198(4323):1264.
Jöbsis FF, Keizer LH, LaManna JC, et al. 1977. Reflectance spectrophotometry of cytochrome aa_3 in vivo. *J Appl Physiol* 43(5):858.
Kelleher JF. 1989. Pulse oximetry. *J Clin Monit* 5(1):37.
Martens H, Nass T. 1989. *Multivariate Calibration.* New York, John Wiley.
Osborne BG, Fearn T, Hindle PH. 1993. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis.* Essex, England, Longman Scientific & Technical.
Payne JP, Severinghaus JW (eds). 1986. *Pulse Oximetry.* New York, Springer-Verlag.
Rabinovitch B, March WF, Adams RL. 1982. Noninvasive glucose monitoring of the aqueous humor of the eye: Part I. Measurement of very small optical rotations. *Diabetes Care* 5(3):254.
Robinson MR, Eaton RP, Haaland DM, et al. 1992. Noninvasive glucose monitoring in diabetic patients: A preliminary evaluation. *Clin Chem* 38(9): 1618.
Severinghaus JW, Astrup PB. 1986. History of blood gas analysis. VI. Oximetry. *J Clin Monit* 2(4):135.

Severinghaus JW, Honda Y. 1987a. History of blood gas analysis. VII. Pulse oximetry. *J Clin Monit* 3(2):135.
Severinghaus JW, Honda Y. 1987b. Pulse oximetry. *Int Anesthesiol Clin* 25(4):205.
Severinghaus JW, Kelleher JF. 1992. Recent developments in pulse oximetry. *Anesthesiology* 76(6): 1018.
Tremper KK, Barker SJ. 1989. Pulse oximetry. *Anesthesiology* 70(1):98.
Wukitsch MW, Petterson MT, Tobler DR, et al. 1988. Pulse oximetry: Analysis of theory, technology, and practice. *J Clin Monit* 4(4):290.
Yoo KM, Alfano RR. 1989. Photon localization in a disordered multilayered system. *Phys Rev B* 39(9):5806.

Further Information

Two collections of papers on pulse oximetry include a book edited by J.P.Payne and J.W.Severinghaus, *Pulse Oximetry* (New York, Springer-Verlag, 1986), and a journal collection—*International Anesthesiology Clinics* [25(4), 1987]. For technical reviews of pulse oximetry, see J.A.Pologe, 1987, Pulse oximetry [*Int Anesthesiol Clin* 25(3):137], Kevin K.Tremper and Steven J.Barker, 1989, Pulse oximetry [*Anesthesiology* 70(1):98], and Michael W.Wukitsch, Michael T.Patterson, David R.Tobler, and coworkers, 1988, Pulse oximetry: Analysis of theory, technology, and practice [*J Clin Monit* 4(4):290].

For a review of practical and clinical applications of pulse oximetry, see the excellent review by Joseph K.Kelleher [1989] and John Severinghaus and Joseph F.Kelleher [1992]. John Severinghaus and Yoshiyuki Honda have written several excellent histories of pulse oximetry [1987*a*, 1987*b*].

For an overview of applied near-infrared spectroscopy, see Donald A.Burns and Emil W.Ciurczak [1992] and B.G.Osborne, T.Fearn, and P.H.Hindle [1993]. For a good overview of multivariate methods, see Harald Martens and Tormod Næs [1989].

31
Medical Instruments and Devices Used in the Home

Bruce R.Bowman
EdenTec Corporation

Edward Schuck
EdenTec Corporation

31.1 Scope of the Market for Home Medical Devices

The market for medical devices used in the home and alternative sites has increased dramatically in the last 10 years and has reached an overall estimated size of more than $1.6 billion [FIND/SVP, 1992]. In the past, hospitals have been thought of as the only places to treat sick patients. But with the major emphasis on reducing healthcare costs, increasing numbers of sicker patients move from hospitals to their homes. Treating sicker patients outside the hospital places additional challenges on medical device design and patient use. Equipment designed for hospital use can usually rely on trained clinical personnel to support the devices. Outside the hospital, the patient and/or family members must be able to use the equipment, requiring these devices to have a different set of design and safety features. This chapter will identify some of the major segments using medical devices in the home and discuss important design considerations associated with home use.

Table 31.1 outlines market segments where devices and products are used to treat patients outside the hospital [FIND/SVP, 1992]. The durable medical equipment market is the most established market providing aids for patients to improve access and mobility. These devices are usually not life supporting or sustaining, but in many cases they can make the difference in allowing a patient to be able to function outside a hospital or nursing or skilled facility. Other market segments listed employ generally more sophisticated solutions to clinical problems. These will be discussed by category of use.

The incontinence and ostomy area of products is one of the largest market segments and is growing in direct relationship to our aging society. Whereas sanitary pads and colostomy bags are not very “high-tech,” well-designed aids can have a tremendous impact on the comfort and independence of these patients. Other solutions to incontinence are technically more sophisticated, such as use of electric stimulation of the sphincter muscles through an implanted device or a miniature stimulator inserted as an anal or vaginal plug to maintain continence [Wall et al., 1993].

0-8493-1813-0/03/$0.00+$1.50

TABLE 31.1 Major Market Segments Outside Hospitals

Market Segment	Estimated Equipment Size (1991)	Device Examples
Durable medical equipment	$373 M*	Specialty beds, wheelchairs, toilet aids, ambulatory aids
Incontinence and ostomy products	$600 M*	Sanitary pads, electrical stimulators, colostomy bags
Respiratory equipment	$180 M*	Oxygen therapy, portable ventilators, nasal CPAP, monitors, apnea monitors
Drug infusion, drug measurement	$300 M	Infusion pumps, access ports, patient-controlled analgesia (PCA), glucose measurement, implantable pumps
Pain control and functional stimulation	$140 M	Transcutaneous electrical nerve stimulation (TENS), functional electrical nerve stimulation (FES)

**Source:* FIND/SVP [1992].

Many forms of equipment are included in the respiratory segment. These devices include those that maintain life support as well as those that monitor patients' respiratory function. These patients, with proper medical support, can function outside the hospital at a significant reduction in cost and increased patient comfort [Pierson, 1994]. One area of this segment, infant apnea monitors, provides parents or caregivers the cardio/respiratory status of an at-risk infant so that intervention (CPR etc.) can be initiated if the baby has a life-threatening event. The infant monitor shown in Fig. 31.1 is an example of a patient monitor designed for home use and will be discussed in more detail later in this chapter. Pulse oximetry monitors are also going home with patients. They are used to measure noninvasively the oxygen level of patients receiving supplemental oxygen or ventilator-dependent patients to determine if they are being properly ventilated.

Portable infusion pumps are an integral part of providing antibiotics, pain management, chemotherapy, and parenteral and enteral nutrition. The pump shown in Fig. 31.2 is an example of technology that allows the patient to move about freely while receiving sometimes lengthy drug therapy. Implantable drug pumps are also available for special long-term therapy needs.

Pain control using electric stimulation in place of drug therapy continues to be an increasing market. The delivery of small electric impulses to block pain is continuing to gain medical acceptance for treatment outside the hospital setting. A different form of electric stimulation called functional electric stimulation

FIGURE 31.1 Infant apnea monitor used in a typical home setting (photo courtesy of EdenTec Corporation).

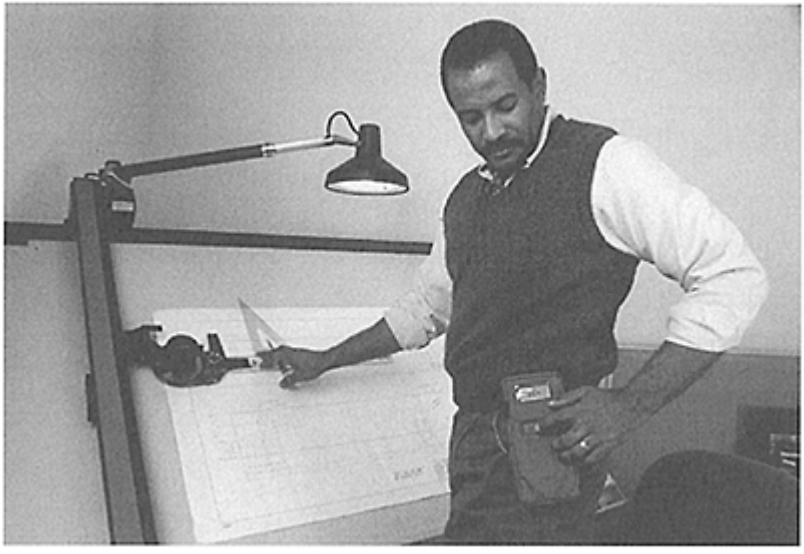


FIGURE 31.2 Portable drug pump used throughout the day (photo courtesy of Pharmacia Deltec Inc.).

(FES) applies short pulses of electric current to the nerves that control weak or paralyzed muscles. This topic is covered as a separate chapter in this book.

Growth of the homecare business has created problems in overall healthcare costs since a corresponding decrease in hospital utilization has not yet occurred. In the future, however, increased homecare will necessarily result in reassessment and downsizing in the corresponding hospital segment. There will be clear areas of growth and areas of consolidation in the new era of healthcare reform. It would appear, however, that homecare has a bright future of continued growth.

31.2 Unique Challenges to the Design and Implementation of High-Tech Homecare Devices

What are some of the unique requirements of devices that could allow more sophisticated equipment to go home with ordinary people of varied educational levels without compromising their care? Even though each type of clinical problem has different requirements for the equipment that must go home with the patient, certain common qualities must be inherent in most devices used in the home. Three areas to consider when equipment is used outside of the hospital are that the device (1) must provide a positive clinical outcome, (2) must be safe and easy to use, and (3) must be user-friendly enough so that it will be used.

The Device Must Provide a Positive Clinical Outcome

Devices cannot be developed any longer just because new technology becomes available. They must solve the problem for which they were intended and make a significant clinical difference in the outcome or management of the patient while saving money. These realities are being driven by those who reimburse for devices, as well as by the FDA as part of the submission for approval to market a new device.

The Device Must Be Safe to Use

Homecare devices may need to be even *more* reliable and even *safer* than hospital devices. We often think of hospitals as having the best quality and most expensive devices that money can buy. In addition to having the best equipment to monitor patients, hospitals have nurses and aids that keep an eye on patients so that equipment problems may be quickly discovered by the staff. A failure in the home may go unnoticed until it is too late. Thus systems for home use really need extra reliability with automatic backup systems and/or early warning signals.

Safety issues can take on a different significance depending on the intended use of the device. Certain safety issues are important regardless of whether the device is a critical device such as an implanted cardiac pacemaker or a noncritical device such as a bed-wetting alarm. No device should be able to cause harm to the patient regardless of how well or poorly it may be performing its intended clinical duties. Devices must be safe when exposed to all the typical environmental conditions to which the device could be

exposed while being operated by the entire range of possible users of varied education and while exposed to siblings and other untrained friends or relatives. For instance, a bedwetting alarm should not cause skin burns under the sensor if a glass of water spills on the control box. This type of safety issue must be addressed even when it significantly affects the final cost to the consumer.

Other safety issues are not obviously differentiated as to being actual safety issues or simply nuisances or inconveniences to the user. It is very important for the designer to properly define these issues; although some safety features can be included with little or no extra cost, other safety features may be very costly to implement. It may be a nuisance for the patient using a TENS pain control stimulator to have the device inadvertently turned off when its on/off switch is bumped while watching TV. In this case, the patient only experiences a momentary cessation of pain control until the unit is turned back on. But it could mean injuries or death to the same patient driving an automobile who becomes startled when his TENS unit inadvertently turns on and he causes an accident.

Reliability issues can also be mere inconveniences or major safety issues. Medical devices should be free of design and materials defects so that they can perform their intended functions reliably. Once again, reliability does not necessarily need to be expensive and often can be obtained with good design. Critical devices, i.e., devices that could cause death or serious injury if they stopped operating properly, may need to have redundant systems for backup, which likely will increase cost.

The Device Must be Designed So That It *Will* Be Used

A great deal of money is being spent in healthcare on devices for patients that end up not being used. There are numerous reasons for this happening including that the wrong device was prescribed for the patient's problem in the first place; the device works, but it has too many false alarms; the device often fails to operate properly; it is cumbersome to use or difficult to operate or too uncomfortable to wear.

Ease of Use

User friendliness is one of the most important features in encouraging a device to be used. Technological sophistication may be just as necessary in areas that allow ease of use as in attaining accuracy and reliability in the device. The key is that the technologic sophistication be transparent to the user so that the device does not intimidate the user. Transparent features such as automatic calibration or automatic sensitivity adjustment may help allow successful use of a device that would otherwise be too complicated.

Notions of what makes a device easy to use, however, need to be thoroughly tested with the patient population intended for the device. Caution needs to be taken in defining what "simple" means to different people. A VCR may be simple to the designer because all features can be programmed with one button, but it may not be simple to users if they have to remember that it takes two long pushes and one short to get into the clock-setting program.

Convenience for the user is also extremely important in encouraging use of a device. Applications that require devices to be portable must certainly be light enough to be carried. Size is almost always important for anything that must fit within the average

household. Either a device must be able to be left in place in the home or it must be easy to set up, clean, and put away. Equipment design can make the difference between the patient appropriately using the equipment or deciding that it is just too much hassle to bother.

Reliability

Users must also have confidence in the reliability of the device being used and must have confidence that if it is not working properly, the device will tell them that something is wrong. Frequent breakdowns or false alarms will result in frustration and ultimately in reduced compliance. Eventually patients will stop using the device altogether. Most often, reliability can be designed into a product with little or no extra cost in manufacturing, and everything that can be done at no cost to enhance reliability should be done. It is very important, however, to understand what level of additional reliability involving extra cost is necessary for product acceptance. Reliability can always be added by duplicated backup systems, but the market or application may not warrant such an approach. Critical devices that are implanted, such as cardiac pacemakers, have much greater reliability requirements, since they involve not only patient frustration but also safety.

Cost Reimbursement

Devices must be paid for before the patient can realize the opportunity to use new, effective equipment. Devices are usually paid for by one of two means. First, they are covered on an American Medical Association Current Procedural Terminology Code (CPT-code), which covers the medical, surgical, and diagnostic services provided by physicians. The CPT-codes are usually priced out by Medicare to establish a baseline reimbursement level. Private carriers usually establish a similar or different level of reimbursement based on regional or other considerations. Gaining new CPT-codes for new devices can take a great deal of time and effort. The second method is to cover the procedure and device under a capitated fee, where the hospital is reimbursed a lump sum for a procedure including the device, hospital, homecare, and physician fees.

Every effort should be made to design devices to be low cost. Device cost is being scrutinized more and more by those who reimburse. It is easy to state, however, that a device needs to be inexpensive. Unfortunately the reality is that healthcare reforms and new regulations by the FDA are making medical devices more costly to develop, to obtain regulatory approvals for [FDA, 1993], and to manufacture.

Professional Medical Service Support

The more technically sophisticated a device is, the more crucial it is that homecare support and education be a part of the program. In fact, in many cases, such support and education are as important as the device itself.

Medical service can be offered by numerous homecare service companies. Typically these companies purchase the equipment instead of the patient, and a monthly fee is charged for use of the equipment along with all the necessary service. The homecare

company then must obtain reimbursement from third-party payers. Some of the services offered by the homecare company include training on how to use the equipment, CPR training, transporting the equipment to the home, servicing/repairing equipment, monthly visits, and providing on-call service 24 hours a day. The homecare provider must also be able to provide feedback to the treating physician on progress of the treatment. This feedback may include how well the equipment is working, the patient's medical status, and compliance of the patient.

31.3 Infant Monitor Example

Many infants are being monitored in the home using apnea monitors because they have been identified with breathing problems [Kelly, 1992]. These include newborn premature babies who have apnea of prematurity [Henderson-Smart, 1992; NIH, 1987], siblings of babies who have died of sudden infant death syndrome (SIDS) [Hunt, 1992; NIH, 1987], or infants who have had an apparent life-threatening episode (ALTE) related to lack of adequate respiration [Kahn et al., 1992; NIH, 1987]. Rather than keeping infants in the hospital for a problem that they may soon outgrow (1 to 6 months), doctors often discharge them from the hospital with an infant apnea monitor that measures the duration of breathing pauses and heart rate and sounds an alarm if either parameter crosses limits prescribed by the doctor.

Infant apnea monitors are among the most sophisticated devices used routinely in the home. These devices utilize microprocessor control, sophisticated breath-direction and artifact rejection firmware algorithms, and internal memory that keeps track of use of the device as well as recording occurrence of events and the physiologic waveforms associated with the events. The memory contents can be downloaded directly to computer or sent via modem remotely where a complete 45-day report can be provided to the referring physician (see Fig. 31.3).

Most apnea monitors measure breathing effort through impedance pneumography. A small (100 to 200 uA) high-frequency (25 to 100 kHz) constant-current train of pulses is applied across the chest between a pair of electrodes. The voltage needed to drive the current is measured, and thereby the effective impedance between the electrodes can be calculated. Impedance across the chest increases as the chest expands and decreases as the chest contracts with each breath. The impedance change with each breath can be as low as 0.2 ohms on top of an electrode base impedance of 2000 ohms, creating some interesting signal-to-noise challenges. Furthermore, motion artifact and blood volume changes in the heart and chest can cause impedance changes of 0.6 ohms or more that can look just like breathing. Through the same pair of electrodes, heart rate is monitored by picking up the electrocardiogram (EGG) [AAMI, 1988].

Because the impedance technique basically measures the motion of the chest, this technique can only be used to monitor central apnea or lack of breathing effort. Another less common apnea in infants called obstructive apnea results when an obstruction of the airway blocks air from flowing in spite of breathing effort. Obstructive apnea cannot be monitored using impedance pneumography [Kelly, 1992].

There is a very broad socioeconomic and educational spectrum of parents or caregivers who may be monitoring their infants with an apnea monitor. This creates an

incredible challenge for the design of the device so that it is easy enough to be used by a variety of caregivers. It also puts special requirements on the homecare service company that must be able to respond to these patients within a matter of minutes, 24 hours a day.

The user-friendly monitor shown in Fig. 31.1 uses a two-button operation, the on/off switch, and a reset switch. The visual alarm indicators are invisible behind a back-lit panel except when an actual alarm occurs. A word describing the alarm then appears. By not showing all nine possible alarm conditions unless an alarm occurs, parent confusion and anxiety is minimized. Numerous safety features are built into the unit, some of which are noticeable but many of which are internal to the operation of the monitor. One useful safety feature is the self-check. When the device is turned on, each alarm LED lights in sequence, and the unit beeps once indicating that the self-check was completed successfully. This gives users the opportunity to confirm that all the alarm visual indicators and the audible indicator are working and provides added confidence for users leaving their baby on the monitor. A dual-level battery alarm gives an early warning that the battery will soon need charging. The weak-battery alarm allows users to reset the monitor and continue monitoring their babies for several more hours before depleting the battery to the charge battery level where the monitor must be attached to the ac battery charger/adapter. This allows parents the freedom to leave their homes for a few hours knowing that their child can continue to be monitored.

A multistage alarm reduces the risk of parents' sleeping through an alarm. Most parents are sleep-deprived with a new baby. Consequently, it can be easy for parents in a nearby room to sleep through a monitor alarm even when the monitor sounds at 85 dB. A three-stage alarm helps to reduce this risk. After 10 seconds of sounding at 1 beep per second, the alarm switches to 3 beeps per second for the next 10 seconds. Finally, if an alarm has not resolved itself after 20 seconds, the alarm switches to 6 beeps per second. Each stage of alarm sounds more intense than the previous one and offers the chance of jolting parents out of even the deepest sleep.

The physician always prescribes what alarm settings should be used by the homecare service company when setting up the monitor. As a newborn matures, these settings may need to be adjusted. Sometimes the parents can be relied upon for making these setting changes. To allow both accessibility to these switches as well as to keep them safe from unauthorized tampering from a helping brother or sister, a special tamper-resistant adjustment procedure is utilized. Two simultaneous actions are required in order to adjust the alarm limit settings. The reset button must be continually pressed on the front of the unit while changing settings on the back of the unit. Heart-rate levels are set in beats per minute, and apnea duration is set in single-second increments. Rather than using easy-to-set pushbutton switches, "penset" switches are used that require a pen or other sharp implement to make the change. If the proper switch adjustment procedure is not followed, the monitor alarms continuously and displays a switch alarm until the settings are returned to their original settings. A similar technique is used for turning the monitor off. The reset button must first be pressed and then the on/off switch turned to the off position. Violation of this procedure will result in a switch alarm.

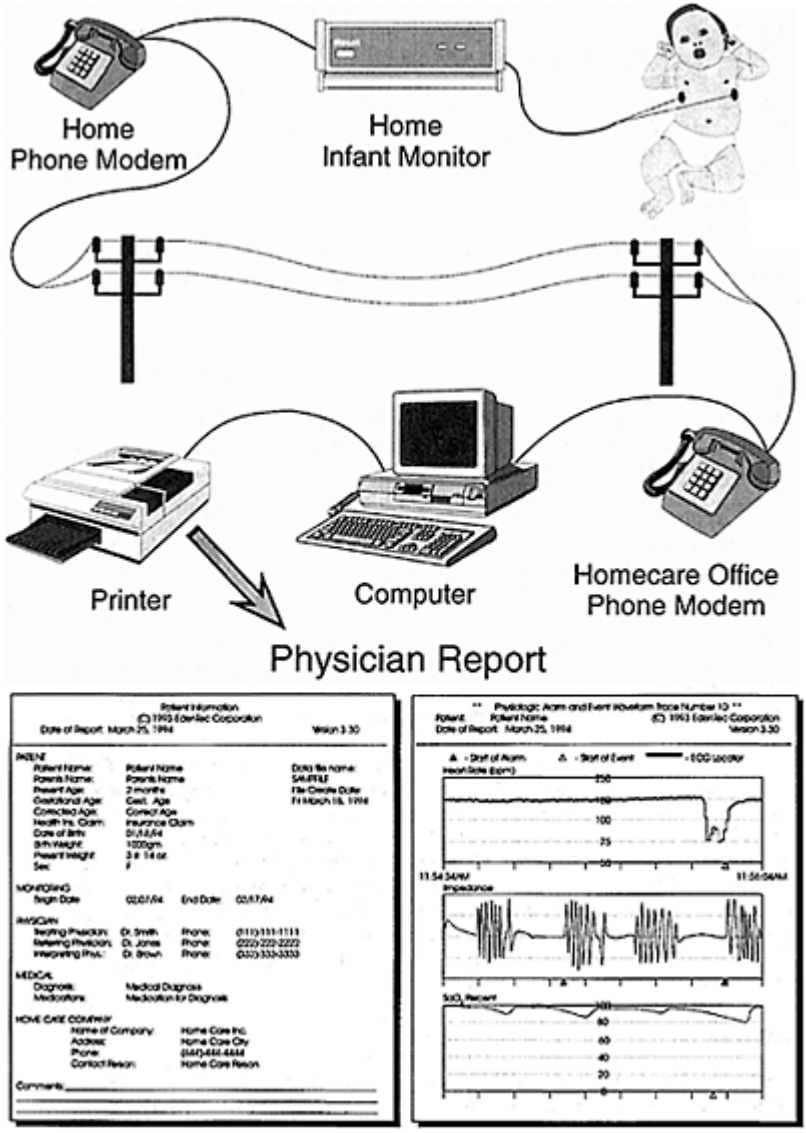


FIGURE 31.3 Infant apnea monitor with memory allows data to be sent by modem to generate physician report (drawing courtesy of EdenTec Corporation).

Other safety features are internal to the monitor and are transparent to the user. The monitor's alarm is designed to be normally on from the moment the device is turned on.

Active circuitry controlled by the microprocessor turns the alarm off when there are no active alarm conditions. If anything hangs up the processor or if any of a number of components fail, the alarm will not turn off and will remain on in a fail-safe mode. This "alarm-on-unless-turned-off" technique is also used in a remote alarm unit for parents with their baby in a distant room. If a wire breakage occurs between the monitor and the remote alarm unit, or a connector pulls loose, or a component fails, the remote alarm no longer is turned off by the monitor and it alarms in a fail-safe condition.

Switches, connectors, and wires are prone to fail. One way to circumvent this potential safety issue is use of switches with a separate line for each possible setting. The monitor continuously polls every switch line of each switch element to check that "exactly" one switch position is making contact. This guards against misreading bad switch elements, a switch inadvertently being set between two positions, or a bad connector or cable. Violation of the "exactly-one-contact condition" results in a switch alarm.

It is difficult to manage an apnea-monitoring program in rural areas where the monitoring family may be a hundred miles or more away from the homecare service company. There are numerous ways to become frustrated with the equipment and stop using the monitor. Therefore, simplicity of use and reliability are important. Storing occurrence of alarms and documenting compliance in internal memory in the monitor help the homecare service company and the remote family cope with the situation. The monitor shown in Fig. 31.1 stores in digital memory the time, date, and duration of (1) each use of the monitor; (2) occurrence of all equipment alarms; and (3) all physiologic alarms including respiratory waveforms, heart rate, and EGG for up to a 45-day period. These data in the form of a report (see Fig. 31.3) can be downloaded to a laptop PC or sent via modem to the homecare service company or directly to the physician.

31.4 Conclusions

Devices that can provide positive patient outcomes with reduced overall cost to the healthcare system while being safe, reliable, and user-friendly will succeed based on pending healthcare changes. Future technology in the areas of sensors, communications, and memory capabilities should continue to increase the potential effectiveness of homecare management programs by using increasingly sophisticated devices. The challenge for the medical device designer is to provide cost-effective, reliable, and easy-to-use solutions that can be readily adopted by the multidisciplinary aspects of homecare medicine while meeting FDA requirements.

Defining Terms

Apnea: Cessation of breathing. Apnea can be classified as **central, obstructive,** or **mixed,** which is a combination.

Apnea of prematurity: Apnea in which the incidence and severity increases with decreasing gestational age attributable to immaturity of the respiratory control system. The incidence has increased due to improved survival rates for very-low-birth-weight premature infants.

Apparent life-threatening episode (ALTE): An episode characterized by a combination of apnea, color change, muscle tone change, choking, or gagging. To the observer it may appear the infant has died.
Capitated fee: A fixed payment for *total* program services versus the more traditional fee for service in which each individual service is charged.
Cardiac pacemaker: A device that electrically stimulates the heart at a certain rate used in absence of normal function of the heart's sino-atrial node.
Central apnea: Apnea secondary to lack of respiratory or diaphragmatic effort.
Chemotherapy: Treatment of disease by chemical agents. Term popularly used when fighting cancer chemically.
Colostomy: The creation of a surgical hole as an alternative opening of the colon.
CPR (cardiopulmonary resuscitation): Artificially replacing heart and respiration function through rhythmic pressure on the chest.
CPT-code (current procedural terminology code): A code used to describe specific procedures/tests developed by the AMA.
Electrocardiogram (ECG): The electric potential recorded across the chest due to depolarization of the heart muscle with each heartbeat.
Enteral nutrition: Chemical nutrition injected intestinally.
Food and Drug Administration (FDA): Federal agency that oversees and regulates foods, drugs, and medical devices.
Functional electrical stimulation (FES): Electric stimulation of peripheral nerves or muscles to gain functional, purposeful control over partially or fully paralyzed muscles.
Incontinence: Loss of voluntary control of the bowel or bladder.
Obstructive apnea: Apnea in which the effort to breath continues but airflow ceases due to obstruction or collapse of the airway.
Ostomy: Surgical procedure that alters the bladder or bowel to eliminate through an artificial passage.
Parenteral nutrition: Chemical nutrition injected subcutaneously, intramuscular, intrasternally, or intravenously.
Sphincter: A band of muscle fibers that constricts or closes an orifice.
Sudden infant death syndrome (SIDS): The sudden death of an infant which is unexplained by history or postmortem exam.
Transcutaneous electrical nerve stimulation (TENS): Electrical stimulation of sensory nerve fibers resulting in control of pain.

References

AAMI. 1988. Association for the Advancement of Medical Instrumentation Technical Information Report. *Apnea Monitoring by Means of Thoracic Impedance Pneumography*, Arlington, Virg.

FDA. November 1993. *Reviewers Guidance for Premarket Notification Submissions* (Draft), Anesthesiology and Respiratory Device Branch, Division of Cardiovascular, Respiratory, and Neurological Devices. Food and Drug Administration. Washington, DC.

FIND/SVP. 1992. *The Market for Home Care Products, a Market Intelligence Report.* New York.

Henderson-Smart DJ. 1992. Apnea of prematurity. In R Beckerman, R Brouillette, C Hunt (eds), *Respiratory Control Disorders in Infants and Children*, pp 161–177, Baltimore, Williams and Wilkins.

Hunt CE. 1992. Sudden infant death syndrome. In R Beckerman, R Brouillette, C Hunt (eds), *Respiratory Control Disorders in Infants and Children,* pp 190–211, Baltimore, Williams and Wilkins.

Kahn A, Rebuffat E, Franco P, et al. 1992. Apparent life-threatening events and apnea of infancy. In R Beckerman, R Brouillette, C Hunt (eds), *Respiratory Control Disorders in Infants and Children,* pp 178–189, Baltimore, Williams and Wilkins.

Kelly DH. 1992. Home monitoring. In R Beckerman, R Brouillette, C Hunt (eds), *Respiratory Control Disorders in Infants and Children,* pp 400–412, Baltimore, Williams and Wilkins.

NIH. 1987. *Infantile Apnea and Home Monitoring Report of NIH Consensus Development Conference,* US Department of Health and Human Services, NIH publication 87–2905.

Pierson DJ. 1994. Controversies in home respiratory care: Conference summary. *Respir Care* 39(4):294.

Wall LL, Norton PA, Dehancey JOL. 1993. *Practical Urology,* Baltimore, Williams and Wilkins.

32
Virtual Instrumentation: Applications in Biomedical Engineering

Eric Rosow
Hartford Hospital and Premise Development Corporation

Joseph Adam
Premise Development Corporation

32.1 Overview

32.1.1 A Revolution—Graphical Programming and Virtual Instrumentation

Over the last decade, the graphical programming revolution has empowered engineers to develop customized systems, the same way the spreadsheet has empowered business managers to analyze financial data. This software technology has resulted in another type of revolution—the virtual instrumentation revolution, which is rapidly changing the instrumentation industry by driving down costs without sacrificing quality.

Virtual Instrumentation can be defined as:

> A layer of software and/or hardware added to a general-purpose computer in such a fashion that users can interact with the computer as though it were their own custom-designed traditional electronic instrument.

Today, computers can serve as the engine for instrumentation. Virtual instruments utilize the open architecture of industry-standard computers to provide the processing, memory, and display capabilities; while the off-the-shelf, inexpensive interface boards plugged into an open bus, standardized communications bus provides the vehicle for the instrument's capabilities. As a result, the open architecture of PCs and workstations allow the functionality of virtual instruments to be user defined. In addition, the processing power of virtual instruments is much greater than stand-alone instruments. This advantage will continue to accelerate due to the rapid technology evolution of PCs and workstations that results from the huge investments made in this industry.

0-8493-1813-0/03/$0.00+$1.50

The major benefits of virtual instrumentation include increased performance and reduced costs. In addition, because the user controls the technology through software, the flexibility of virtual instrumentation is unmatched by traditional instrumentation. The modular, hierarchical programming environment of virtual instrumentation is inherently reusable and reconfigurable.

32.2 Virtual Instrumentation and Biomedical Engineering

Virtual Instrumentation applications have encompassed nearly every industry including the telecommunications, automotive, semiconductor, and biomedical industries. In the fields of healthcare and biomedical engineering, virtual instrumentation has empowered developers and end-users to conceive of, develop, and implement a wide variety of research-based biomedical applications and executive information tools. These applications fall into several categories including: clinical research, equipment testing and quality assurance, data management, and performance improvement.

In a collaborative approach, physicians, researchers, and biomedical and software engineers at Hartford Hospital (Hartford, CT) and Premise Development Corporation (Avon, CT) have developed various data-acquisition and analysis systems that successfully integrate virtual instrumentation principles in a wide variety of environments. These include:

- "The EndoTester™", a patented quality assurance system for fiberoptic endoscopes
- a noninvasive pulmonary diffusion and cardiac output measurement system
- a cardiovascular pressure-dimension analysis system
- "BioBench™" a powerful turnkey application for physiological data acquisition and analysis
- "PIVIT™", a performance indicator virtual instrument toolkit to manage and forecast financial data
- a "virtual intelligence program" to manage the discrete components within the continuum of care
- "BabySave™", an analysis and feedback system for apnea interruption via vibratory stimulation

This chapter will describe several of these applications and describe how they have allowed clinicians and researchers to gain new insights, discover relationships that may not have been obvious, and test and model hypotheses based on acquired data sets. In some cases, these applications have been developed into commercial products to address test and measurement needs at other healthcare facilities throughout the world.

Example 1: BioBench™—A Virtual Instrument Application for Data Acquisition and Analysis of Physiological Signals

The biomedical industry is an industry that relies heavily on the ability to acquire, analyze, and display large quantities of data. Whether researching disease mechanisms and treatments by monitoring and storing physiological signals, researching the effects of various drugs interactions, or teaching students in labs where students study physiological signs and symptoms, it was clear that there existed a strong demand for a flexible, easy-to-use, and cost-effective tool. In a collaborative approach, biomedical engineers, software engineers and clinicians, and researchers created a suite of virtual instruments called BioBench™.

BioBench™ (National Instruments, Austin, TX) is a new software application designed for physiological data acquisition and analysis. It was built with Lab VIEW™, the world's leading software development environment for data acquisition, analysis, and presentation.[1] Coupled with National Instruments data acquisition (DAQ) boards, BioBench integrates the PC with data acquisition for the life-sciences market.

Many biologists and physiologists have made major investments over time in data-acquisition hardware built before the advent of modern PCs. While these scientists cannot afford to throw out their investment in this equipment, they recognize that computers and the concept of virtual instrumentation yield tremendous benefits in terms of data analysis, storage, and presentation. In many

[1] BioBench™ was developed for National Instruments (Austin, TX) by Premise Development Corporation (Avon, CT).

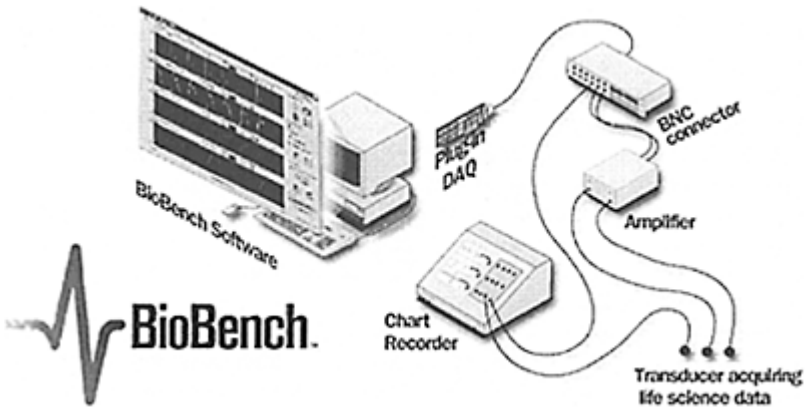


FIGURE 32.1 A typical biomedical application using BioBench (courtesy of National Instruments).

cases, traditional medical instrumentation may be too expensive to acquire and/or maintain. As a result, researchers and scientists are opting to create their own PC-based data monitoring systems in the form of virtual instruments.

Other life scientists, who are just beginning to assemble laboratory equipment, face the daunting task of selecting hardware and software needed for their application. Many manufacturers for the life sciences field focus their efforts on the acquisition of raw signals and converting these signals into measurable linear voltages. They do not concentrate on digitizing signals or the analysis and display of data on the PC. BioBench™ is a low-cost turnkey package that requires no programming. BioBench is compatible with any isolation amplifier or monitoring instrument that provides an analog output signal. The user can acquire and analyze data immediately because BioBench automatically recognizes and controls the National Instruments DAQ hardware, minimizing configuration headaches.

Some of the advantages of PC-based data monitoring include:

- Easy-to-use-software applications
- Large memory and the PCI bus
- Powerful processing capabilities
- Simplified customization and development
- More data storage and faster data transfer
- More efficient data analysis

Figure 32.1 illustrates a typical setup of a data acquisition experiment using BioBench. BioBench also features pull-down menus through which the user can configure devices. Therefore, those who have made large capital investments can easily migrate their existing equipment into the computer age. Integrating a combination of old and new physiological instruments from a variety of manufacturers is an important and straightforward procedure. In fact, within the clinical and research setting, it is a common requirement to be able to acquire multiple physiological signals from a variety of medical devices and instruments that do not necessarily communicate with each other. Often times, this situation is compounded by the fact that end-users would like to be able to view and analyze an entire waveform and not just an average value. In order to accomplish this, the end-user must acquire multiple channels of data at a relatively high sampling rate and have the ability to manage many large data files. BioBench can collect up to 16 channels simultaneously at a sampling rate of 1000 Hz per channel. Files are stored in an efficient binary format which significantly reduces the amount of hard disk and memory requirements of the PC. During data acquisition, a number of features are available to the end-user. These features include:

Data Logging: Logging can be enabled prior to or during an acquisition. The application will either prompt the user for a descriptive filename or it can be configured to automatically assign a filename for each acquisition. Turning the data logging option on and off creates a log data event record that can be inspected in any of the analysis views of BioBench.

Event Logging: The capacity to associate and recognize user commands associated with a data file may be of significant value. BioBench has been designed to provide this capability by automatically logging user-defined events, stimulus events, and file logging events. With user-defined events, the user can easily enter and associate date and time-stamped notes with user actions or specific subsets of data. Stimulus events are also data and time-stamped and provide the user information about whether a stimulus has been turned on or off. File logging events note when data has been logged to disk. All of these

types of events are stored with the raw data when logging data to file and they can be searched for when analyzing data.

Alarming: To alert the user about specific data values and thresholds, BioBench incorporates user-defined alarms for each signal that is displayed. Alarms appear on the user interface during data acquisition and notify the user that an alarm condition has occurred.

Figure 32.2 is an example of the data acquisition mode of BioBench. Once data has been acquired, BioBench can employ a wide array of easy-to-use analysis features. The user has the choice of importing recently acquired data or opening a data file that had been previously acquired for comparison or teaching purposes. Once a data set has been selected and opened, BioBench allows the user to simply select and highlight a region of interest and choose the analysis options to perform a specific routine.

BioBench implements a wide array of scalar and array analyses. For example, scalar analysis tools will determine the minimum, maximum, mean, integral, and slope of a selected data set, while the array analysis tools can employ fast fourier transforms (FFTs), peak detection, histograms, and X versus Y plots.

The ability to compare multiple data files is very important in analysis and BioBench allows the user to open an unlimited number of data files for simultaneous comparison and analysis. All data files can be scanned using BioBench's search tools in which the user can search for particular events that are

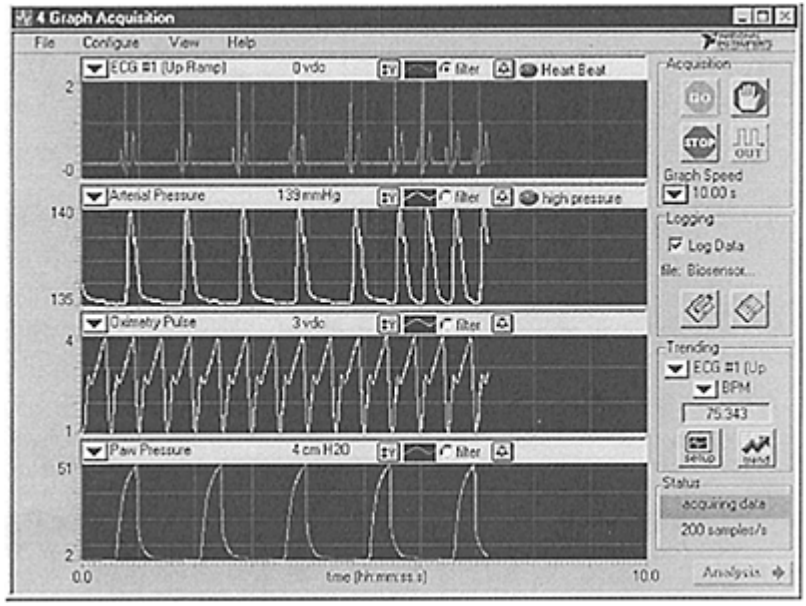


FIGURE 32.2 BioBench acquisition mode with alarms enabled.

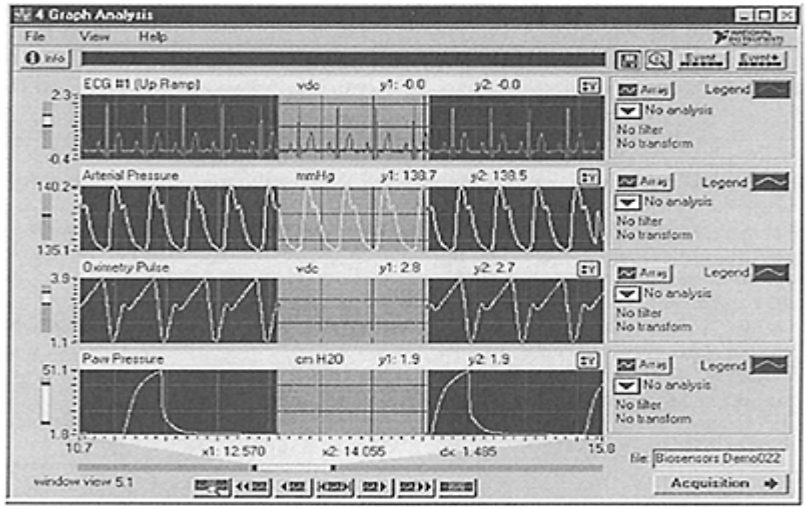


FIGURE 32.3 BioBench analysis mode.

associated with areas of interest. In addition, BioBench allows the user to employ filters and transformations to their data sets and all logged data can be easily exported to a spreadsheet or database for further analysis. Finally, any signal acquired with BioBench can be played back, thus taking lab experience into the classroom. Figure 32.3 illustrates the analysis features of BioBench.

Example 2: A Cardiovascular Pressure-Dimension Analysis System

Introduction

The intrinsic contractility of the heart muscle (myocardium) is the single most important determinant of prognosis in virtually all diseases affecting the heart (e.g., coronary artery disease, valvular heart disease, and cardiomyopathy). Furthermore, it is clinically important to be able to evaluate and track myocardial function in other situations, including chemotherapy (where cardiac dysfunction may be a side effect of treatment) and liver disease (where cardiac dysfunction may complicate the disease).

The most commonly used measure of cardiac performance is the ejection fraction. Although it does provide some measure of intrinsic myocardial performance, it is also heavily influenced by other factors such as heart rate and loading conditions (i.e., the amount of blood returning to the heart and the pressure against which the heart ejects blood).

Better indices of myocardial function based on the relationship between pressure and volume throughout the cardiac cycle (pressure-volume loops) exist. However, these methods have been limited because they require the ability to track ventricular volume continuously during rapidly changing loading conditions. While there are many

techniques to measure volume under steady state situations, or at end-diastole and end-systole (the basis of ejection fraction determinations), few have the potential to record volume during changing loading conditions.

Echocardiography can provide online images of the heart with high temporal resolution (typically 30 frames per second). Since echocardiography is radiation-free and has no identifiable toxicity, it is ideally suited to pressure-volume analyses. Until recently however, its use for this purpose has been limited by the need for manual tracing of the endocardial borders, an extremely tedious and time-consuming endeavor.

The System

Biomedical and software engineers at Premise Development Corporation (Avon, CT), in collaboration with physicians and researchers at Hartford Hospital, have developed a sophisticated research application called the "Cardiovascular Pressure-Dimension Analysis (CPDA) System." The CPDA system acquires echocardiographic volume and area information from the acoustic quantification (AQ) port, in conjunction with vetricular pressure(s) and ECG signals to rapidly perform pressure-volume and pressure-area analyses. This fully automated system allows cardiologists and researchers to perform online pressure-dimension and stroke work analyses during routine cardiac catheterizations and open-heart surgery. The system has been designed to work with standard computer hardware. Analog signals for ECG, pressure, and area/volume (AQ) are connected to a standard BNC terminal board. Automated calibration routines ensure that each signal is properly scaled and allows the user to immediately collect and analyze pressure-dimension relationships.

The CPDA can acquire up to 16 channels of data simultaneously. Typically, only three physiological parameters, ECG, pressure, and the AQ signals are collected using standard data acquisition hardware. In addition, the software is capable of running on multiple operating systems including Macintosh, Windows 95/98/NT, and Solaris. The CPDA also takes advantage of the latest hardware developments and form factors and can be used with either a desktop or a laptop computer.

The development of an automated, online method of tracing endocardial borders (Hewlett Packard's AQ Technology) (Hewlett-Packard Medical Products Group, Andover, MA) has provided a method for rapid online area and volume determinations. Figure 32.4 illustrates this AQ signal from a Hewlett Packard Sonos Ultrasound Machine. This signal is available as an analog voltage (−1 to+1 volts) through the Sonos Dataport option (BNC connector).

Data Acquisition and Analysis

Upon launching this application, the user is presented with a dialog box that reviews the license agreement and limited warranty. Next, the Main Menu is displayed, allowing the user to select from one of six options as shown in Fig. 32.5.

Clinical Significance

Several important relationships can be derived from this system. Specifically, a parameter called the *End-Systolic Pressure-Volume Relationship (ESPVR)* describes the line of best fit through the peak-ratio (maximum pressure with respect to minimum volume) coordinates from a series of pressure-volume

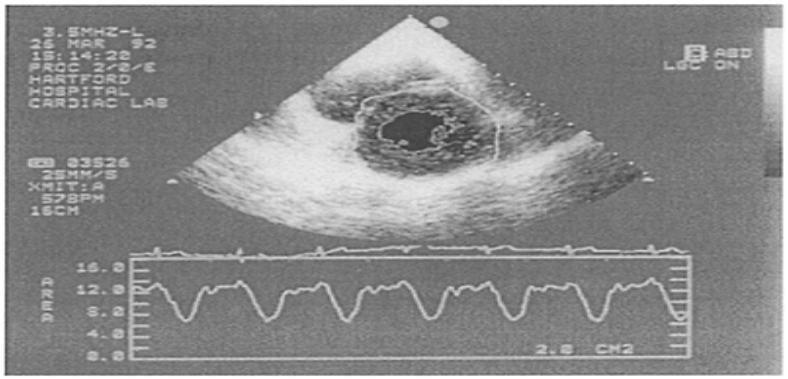


FIGURE 32.4 The Acoustic Quantification (AQ) signal (Hewlett Packard).

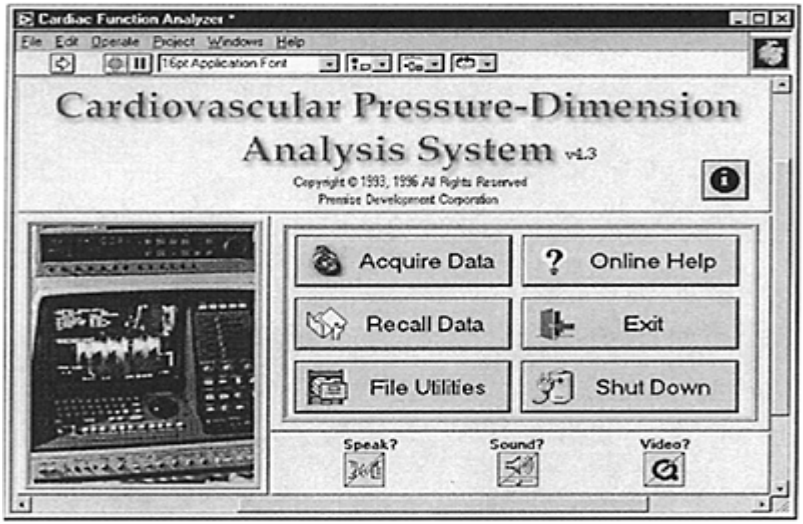


FIGURE 32.5 Cardiovascular pressure-dimension analysis main menu.

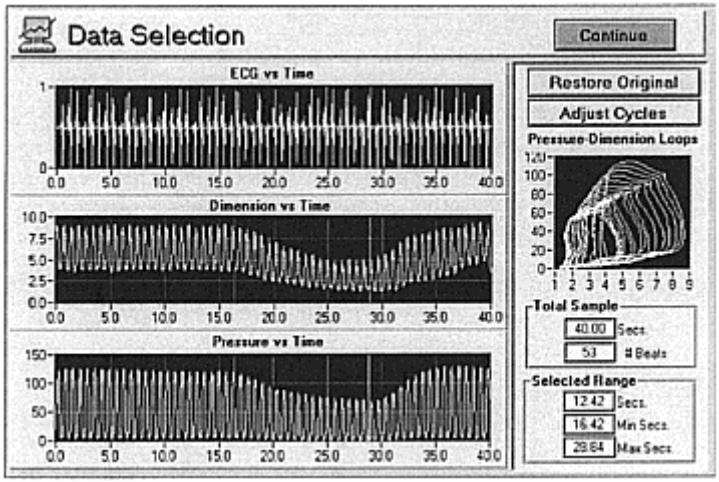


FIGURE 32.6 The data selection front panel.

loops generated under varying loading conditions. The slope of this line has been shown to be a sensitive index of myocardial contractility that is independent of loading conditions. In addition, several other analyses, including *time varying elastance (E_{max})* and *stroke work*, are calculated. Time-varying elastance is measured by determining the maximum slope of a regression line through a series of isochronic pressure-volume coordinates. Stroke work is calculated by quantifying the area of each pressure-volume loop. Statistical parameters are also calculated and displayed for each set of data. Figure 32.7 illustrates the pressure-dimension loops and each of the calculated parameters along with the various analysis options. Finally, the user has the ability to export data sets into spreadsheet and database files and export graphs and indicators into third-party presentation software packages such as Microsoft PowerPoint®.

32.3 Summary

Virtual Instrumentation allows the development and implementation of innovative and cost-effective biomedical applications and information management solutions. As the healthcare industry continues to respond to the growing trends of managed care and capitation, it is imperative for clinically useful, cost-effective technologies to be developed and utilized. As application needs will surely continue to change, virtual instrumentation systems will continue to offer users flexible and powerful solutions without requiring new equipment or traditional instruments.

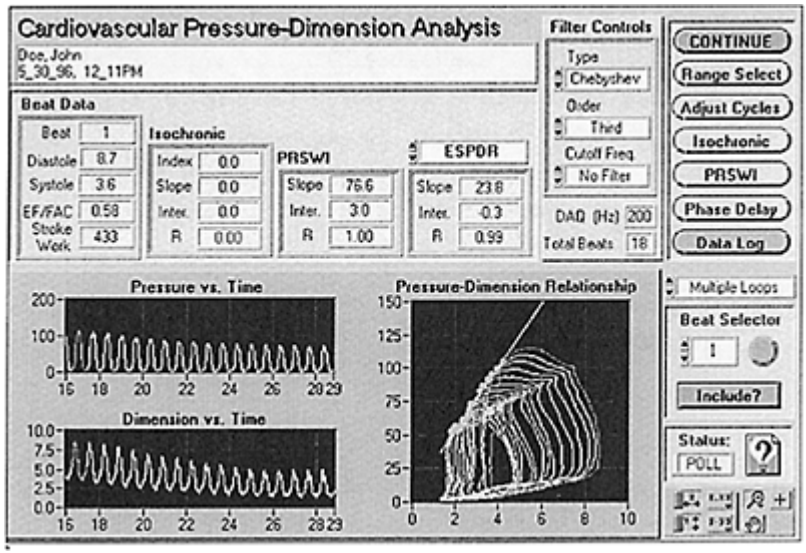


FIGURE 32.7 The cardiac cycle analysis front panel.

References

1. Fisher, J.P., Mikan, J.S., Rosow, E., Nagle, J., Fram, D.B., Maffucci, L.M., McKay, R.G., and Gillam, L.D., Pressure-dimension analysis of regional left ventricular performance using echocardiographic automatic boundary detection: Validation in an animal model of inotropic modulation, *J. Am. College of Cardiol*, 19(3), 262A, 1992.
2. Fisher, J.P., McKay, R.G., Mikan, J.S., Rosow, E., Nagle, J., Mitchel, J.F., Kiernan, F.J., Hirst, J.A., Primiano, C.A., Fram, D.B., Gillam, L.D., Hartford Hospital and University of Connecticut, Hartford, CT, *Human Left Ventricular Pressure-Area and Pressure-Volume Analysis Using Echocardiographic Automatic Boundary Detection*, 65th Scientific Session of the American Heart Association (11/92).
3. Fisher, J.P., Mitchel, J.F., Rosow, E., Mikan, J.S., Nagle, J., Kiernan, F.J., Hirst, J.A., Primiano, Gillam, L.D., Hartford Hospital and University of Connecticut, Hartford, CT, *Evaluation of Left Ventricular Diastolic Pressure-Area Relations with Echocardiographic Automatic Boundary Detection*, 65th Scientific Session of the American Heart Association (11/92).
4. Fisher, J.P., McKay, R.G., Mikan, J.S., Rosow, E., Nagle, J., Mitchel, J.F., Fram, D.B., Gillam, L.D., Hartford Hospital, *A Comparison of Echocardiographic Methods of Evaluating Regional LV Systolic Function: Fractional Area Change vs. the End-Systolic Pressure-Area Relation*, 65th Scientific Session of the American Heart Association (11/92).
5. Fisher, J.P., McKay, R.G., Rosow, E. Mikan, J. Nagle, J. Hirst, J.A., Fram, D.B., Gillam, L.D., Online derivation of human left ventricular pressure-volume loops and load independent indices of contractility using echocardiography with automatic boundary detection—A clinical reality, *Circulation* 88, I-304, 1993.
6. Fisher, J.P., Chen, C., Krupowies, N., Li Li, Kiernan, F.J., Fram, D.B., Rosow, E., Adam, J., Gillam, L.D., Comparison of mechanical and pharmacologic methods of altering loading

conditions to determine end-systolic indices of left ventricle function, *Circulation* 90 (II), 1–494, 1994.
7. Fisher, JP, Martin, J., Day FP, Rosow, E., Adam J, Chen C, Gillam LD. Validation of a less invasive method for determining preload recruitable stroke work derived with echocardiographic automatic boundary detection, *Circulation* 92, 1–278, 1995.
8. Fontes ML. Adam J, Rosow E, Mathew J, DeGraff AC. Non-invasive cardiopulmonary function assessment system, *J Clin Monit* 13, 413, 1997.
9. Johnson, GW, *LabVIEW Graphical Programming: Practical Applications in Instrumentation and Control*, 2nd ed., McGraw-Hill, New York, 1997.
10. Mathew JP, Adam J, Rosow E, Fontes ML, Davis L, Barash PG, Gillam L., Cardiovascular pressure-dimension analysis system, *J Clin Monit* 13, 423, 1997.
11. *National Instruments 1999 Measurement and Automation Catalog*, National Instruments, Austin, TX.
12. Rosow, E. *Technology and Sensors*, Presented at the United States Olympic Committee's Sports Equipment and Technology Conference, Colorado Springs, CO; November 19–23, 1992.
13. Rosow, E. *Biomedical Applications using Lab VIEW*, Presented at the New England Society of Clinical Engineering, Sturbridge, MA; November 14, 1993.
14. Rosow, E., Adam, J.S., Nagle, J., Fisher, J.P. and Gillam, L.D.; Premise Development Corporation, Avon, CT and Hartford Hospital, Hartford, CT, *A Cardiac Analysis System: Lab VIEW and Acoustic Quantification (AQ)*, Presented at the Association for the Advancement of Medical Instrumentation in Washington, D.C., May 21–25, 1994.

Appendix 2
Historical Perspective 2: Recording of Action Potentials

Leslie A.Geddes
Purdue University

Nerve Action Potential

The question for the form and nature of the nerve action potential used instruments that we would now call primitive. Yet, in skilled hands, these instruments led directly to discovery of the code used by the nervous system to transmit information. In fact, the code was discovered before the true form of the nerve action potential was known.

Using a slow-speed galvanometer ballistically to measure time [Hoff and Geddes, 1960], Helmholtz [1850, 1851, 1853] measured the velocity of the frog sciatic nerve action potential and determined it to be 30 m/s. This was far below the speed of electricity and caused much controversy among physiologists of the day. Using the rheotome, a slowly responding galvanometer, and an induction coil stimulator [Geddes et al., 1989], Bernstein [1868] reconstructed the action potential of the frog sciatic nerve [Hoff and Geddes, 1957]. The sampling time was 0.3 ms, and the action potential that he obtained is shown in Fig. A2.1. Not only did Bernstein chart the time course of this 0.5626- to 0.8041-ms action potential, he measured its propagation velocity, obtaining an average of 28.718 m/s, agreeing with the speed obtained by Helmholtz. Thus, with primitive electromechanical instruments, the time course and velocity of the nerve action potential were determined accurately.

When the capillary electrometer appeared in 1876, it was thought that it might be possible to use it to record the nerve action potential. Many attempts were made with variable results because the concept of response time was not fully appreciated. Gotch and Horsley [1888] in the United Kingdom stimulated peripheral nerves in the cat using

induction-coil shocks, and with one recording electrode over intact nerve and one in an area of injury, they recorded the action potential. The type of record that they obtained is shown in Fig. A2.2*a*. Note that the response (action potential) was in the same direction for the break (b) and make (m) shocks. Recall that the break shock from an induction coil is stronger than the make shock. They knew that the make and break shocks were of opposite polarity and proved it by recording them, as shown in Fig. A2.2*b*. After a long discussion about the nerve response being always in the same direction, irrespective of the polarity of the stimulus, they concluded that they had recorded single-action potentials in response to single-induction-coil stimuli. They wrote:

> There is thus no doubt that the movement [of the mercury contour] that we obtained and photographed was due to the electromotive change, which was due to the electromotive change, which accompanies the propagation of an excitatory state along the mammalian nerve when this state is evoked by the application of a single stimulus.

0-8493-1813-0/03/$0.00+$1.50

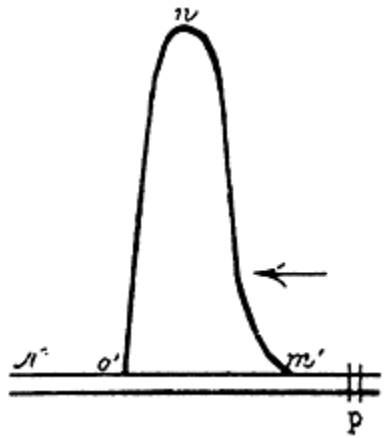


FIGURE A2.1 Bernstein's reconstruction of the action potential of the frog sciatic nerve. From his experiments, the mean duration was 0.6833 ms. This reconstruction was made with a slow-speed galvanometer and the rheotome. (From Bernstein, 1868.)

Like Gotch and Horsley [1888], others who attempted recording nerve action potentials with the capillary electrometer were rewarded with a small-amplitude record. When

Einthoven's string galvanometer appeared in 1903, it was also used to record nerve action potentials with the same result [Einthoven, 1903]. In 1907 and 1908, De Forest patented the triode (audion), and the vacuum-tube amplifier could be constructed. Therefore it appeared logical to use a triode amplifier to enlarge nerve action potentials and display them with the string galvanometer and capillary electrometer.

The first to put the vacuum-tube amplifier to work in electro-physiology were Forbes and Thacher [1920], who coupled a triode to a string galvanometer and recorded frog nerve and human muscle action potentials. Their paper is essentially a tutorial that describes three ways of coupling the triode to the string galvanometer so that the delicate string would not be damaged. The first method placed the vacuum tube in one arm of a Wheatstone bridge. The string galvanometer was in the detector position, and a resistor constituted the arm adjacent to the triode. The second method employed transformer coupling, and the third method employed capacitor *(C)* coupling, as shown in Fig. A2.3*a*. Of the three methods, the capacitive-coupling method was preferred. Figure A2.3*b* is a record obtained by Forbes and Thacher showing the amplified and directly recorded frog sciatic nerve action potential. The timing signal (bottom) is 100 Hz.

Desirous of displaying repetitive nerve action potentials with the string galvanometer, Gasser and Newcomer [1921] employed a two-stage resistance-capacity-coupled amplifier connected to the string galvanometer. They chose the phrenic nerve as the object of their study because of its spontaneous periodic activity in causing the diaphragm to contract tetanically and produce inspiration. The frequency of the phrenic nerve action potentials ranged from 71 to 105 per second, noting that their amplitude appeared to be largest at the peak of inspiration. They also reported a one-to-one correspondence between phrenic nerve and diaphragm action potentials. Their important observations were later to become recognized as the two ways that intensity is signaled in the nervous system.

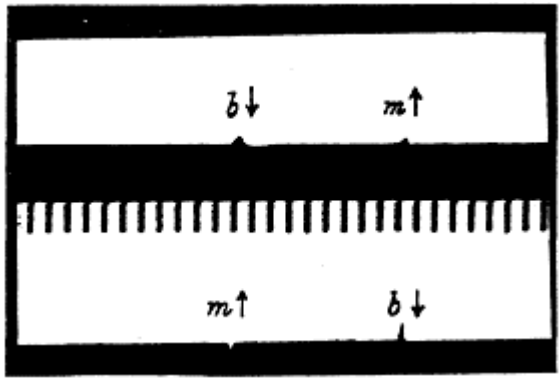


FIGURE A2.2 The first nerve action potentials in response to single stimuli, recorded with the capillary electrometer in response to a "break" (b) and "make" (m) induction-oil

stimulus *(top)*. *(Bottom)* The capillary electrometer record of the break (b) and make (m) stimuli from the induction coil; note that they are of opposite polarity and that of the nerve responses *(top)* are of the same polarity. (From Gotch and Horsley, 1888).

The addition of amplification to the string galvanometer did not improve its ability to respond to rapidly rising, short-duration action potentials. It required the cathode-ray tube and a multistage amplifier to solve this problem. Gasser and Erlanger [1922] were the first to show the form of the nerve action potential recorded extracellularly using the cathode-ray tube and triode amplifier that they had previously developed for use with the string galvanometer. Not only did they achieve their goal, but they made the fundamental discovery that nerve propagation velocity was proportional to nerve fiber diameter.

Although rapidly responding, the cathode-ray tube is quite insensitive, requiring about 20 to 50 V for a 1-cm deflection of the spot made by the electron beam on the face of the tube. Therefore, considerable amplification was needed to display the millivolt nerve action potentials detected with electrodes on the surface of a nerve trunk. Gasser

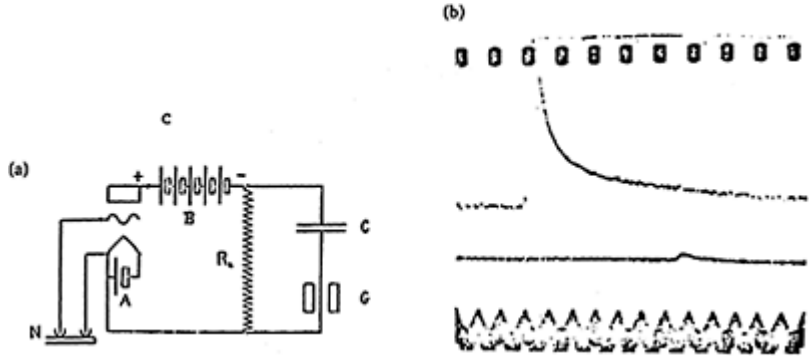


FIGURE A2.3 Use of the capacitively coupled (C) triode to enlarge action potentials from a nerve *(N)* and display them on the string galvanometer *(G)*. *(b) (upper)* An amplified action potential and *(lower)* the unamplified action potential. At the bottom is a record from a 100-Hz tuning fork [Forbes and Thacher 1920].

and Erlanger [1922] built a three-stage amplifier to display nerve action potentials on the cathode-ray tube screen. Figure A2.4 is the circuit diagram of their equipment. The cathode-ray tube was of the low-voltage (300) type and was provided by J.B.Johnson and E.B.Craft of the Western Electric Company. The tube contained a little argon gas. The fluorescent screen was green with a long persistence (5 to 10 s). The deflection sensitivity was 10 to 20 V/cm.

Before describing the Gasser and Erlanger Nobel prize-winning research, it is of interest to examine how the nerve was stimulated and the action potentials were recorded. Figure A2.4 shows the nerve *(N)* on the left with the stimulating electrodes *(T)* and the recording electrodes *(E)*. The nerve was crushed under the electrode connected to the grid *(G)* of the first vacuum-tube amplifier. Therefore, the nerve action potential appeared under the electrode on the uninjured nerve. Three stages of amplification were used to enlarge the nerve action potential 7000 to 8000 times, which caused the beam of the cathode-ray tube to be deflected vertically.

On the right of Fig. A2.4 is a rheotome that (1) delivered the stimulus to an induction coil *(PS)* connected to the nerve-stimulating electrodes *(T)*, (2) started the cathode-ray tube beam moving across the face of the tube to provide a time axis by starting the 1-mF (nowadays μF) capacitor to charge (the voltage on this capacitor was connected to the horizontal deflecting plates in the cathode-ray tube), and (3) later discharged the 1-μF capacitor so that the cycle could be restarted. In this way, the stimulus was delivered to the nerve at the instant when the spot started horizontally across the face of the cathode-ray tube. Typically, 30 stimuli per second were delivered to produce a clear (standing wave) action potential on the cathode-ray tube. Commenting on the action potential recorded from frog sciatic nerve, Gasser and Erlanger [1922] wrote:

> The action current has a gradual ascent, a steep, smooth anacrotic limb [rising phase] and a more gradual catacrotic limb [falling phase]. The latter, like the former, shows a period of great initial acceleration so that the crest is situated near the anacrotic side. In frog nerve and some mammalian nerves, there are secondary waves on the catacrotic limb n. Suggestions are made as to the cause of these waves.

Soon an explanation was given for the secondary waves; it came in 1924 when Erlanger and Gasser [1924] published their classic report, which will now be described.

Having no camera, Gasser and Erlanger either placed tissue paper on the cathode-ray tube face and traced the waveform with a pencil or pressed photographic paper against the face of the tube and obtained a contact print. Figure A2.5 is such an illustration of the action potential of the bullfrog sciatic nerve.

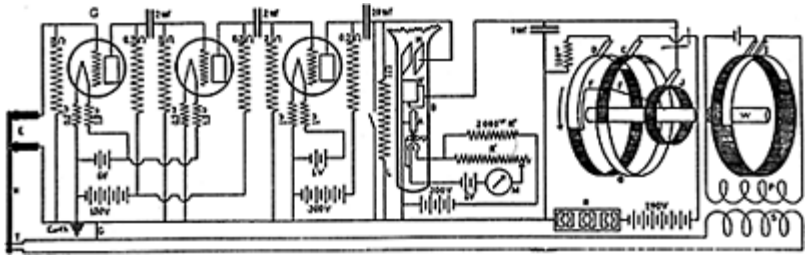


FIGURE A2.4 Circuit diagram of the three-stage amplifier, cathode-ray tube, and stimulator used by Gasser and Erlanger [1922] to record nerve action potentials.

FIGURE A2.5 Contact print from the cathode-ray tube showing the action potential of the bullfrog sciatic nerve recorded by Erlanger and Gasser [1924].

Erlanger and Gasser knew that their oscilloscope time base was exponential and corrected the oscillograms accordingly. A favorite method employed transilluminating a contact print and tracing the action potential on the back with a pencil to obtain a positive image that was then replotted on semilogarthmic paper. This step was essential because they desired a true temporal display of the compound action potential from a nerve trunk. By

varying the distance over which the action potential was propagated, they were able to reveal that the nerve trunk contained groups of fibers that propagated with different velocities. Commenting on their results, they stated:

> Each of the waves of these compound action currents as it progresses along the nerve changes its form just as does the simple action current in the phrenic nerve. It is suggested that these changes are due, in part, at least, to slight differences in the propagation rate in individual, or in many small groups of fibers, whose action currents therefore get slightly out of phase as they progress.

Finally, Erlanger and Gasser [1937] summarized their work in a monograph entitled *Electrical Signs of Nervous Activity*. In it they not only showed how the compound action potential of nerve depends on the velocities of propagation of the different bundles of nerve fibers in a trunk but also provided histograms (fiber maps) of the diameters of different-sized fibers in a nerve trunk. Figure A2.6 shows the action potentials of the A, B, and C fibers. For this pioneering work Erlanger and Gasser received the Nobel prize in physiology and medicine in 1944.

Meanwhile, in the United Kingdom, Adrian was conducting experiments that showed that a nerve fiber responds in an all-or-none manner and that intensity is signaled in a single nerve fiber by the frequency of action potentials. In addition, he showed that intensity is also signaled by the number of nerve fibers carrying messages. These remarkable discoveries were made with the capillary electrometer, which even then was considered to be a primitive instrument. Adrian [1928] defended his use of the capillary electrometer in the following way:

> The ideal instrument for recording nerve action currents is undoubtedly the cathode-ray oscillograph devised by Erlanger and Gasser, for in this, the moving system is a stream of cathode rays, the inertia of which is completely negligible. At present, however, the intensity of the illumination from the ray is far too small to allow photographs to be made from a single excursion, and similar excursions must be repeated many times over before the plate or the eye is affected. As a result, the cathode-ray oscillograph can only be used in experiments where the same sequence of action currents can be repeated over and over again, and it is not suitable for recording an irregular series of action currents such as are produced by the activity of the central nervous system. Another instrument in which the

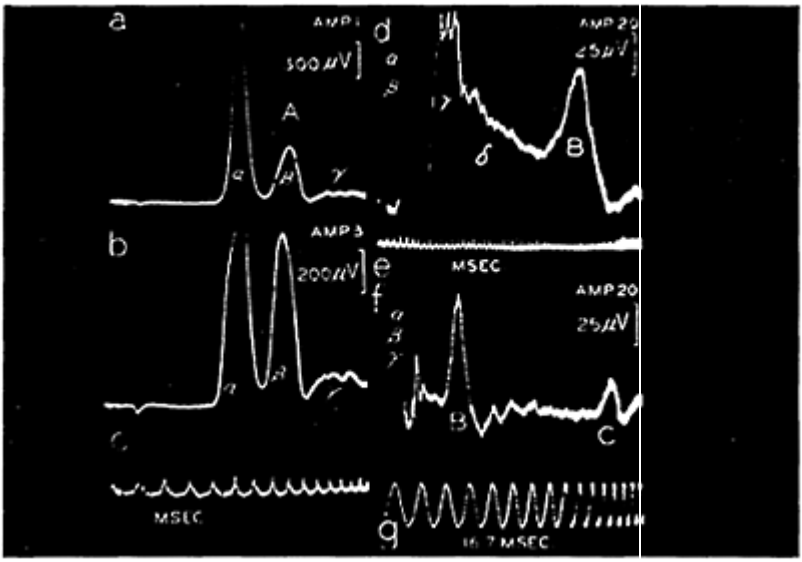


FIGURE A2.6 Action potential components of the A, B, and C fibers of nerve. (From Erlanger J, Gasser HS. *Electrical Signs of Nervous Activity*. Philadelphia, University of Pennsylvania Press, 1937. With permission.)

> inertia factor is extremely small is the capillary electrometer. This has fallen out of favor with the majority of physiologists because its records need analysis and because of its low sensitivity compared with that of the string galvanometer. These objections have now become of little importance. With the advent of reliable valve [vacuum-tube] amplifiers, a low sensitivity in the recording instrument is no drawback at all, and the analysis of capillary electrometer records can be made in a few moments by the machine designed by Keith Lucas. As will be seen, the combination of value amplifier and [capillary] electrometer gives us an instrument of such range and precision that it promises access to fields of investigation, which are as yet almost unexplored.

In Adrian's first studies [1928] he used the three-valve (vacuum-tube) amplifier shown in Fig. A2.7 to record action potentials in different nerve fibers. In his next studies with Bronk [Adrian & Bronk, 1928], he used a single-stage amplifier (Fig. A2.8) that drove a capillary electrometer across that could be connected to earphones or another amplifier

(B) that drove a loudspeaker *(LS)* to enable listening to the nerve action potentials. Adrian stated:

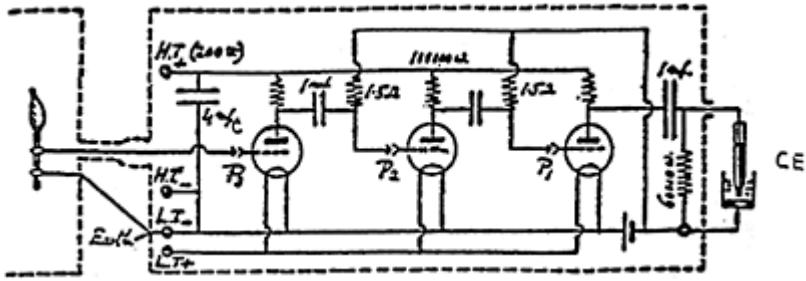


FIGURE A2.7 The three-stage valve (vacuum-tube) amplifier and capillary electrometer (CE) used by Adrian to record action potentials in afferent nerves. The dashed lines represent shielding. (Redrawn from [Adrian, 1926].)

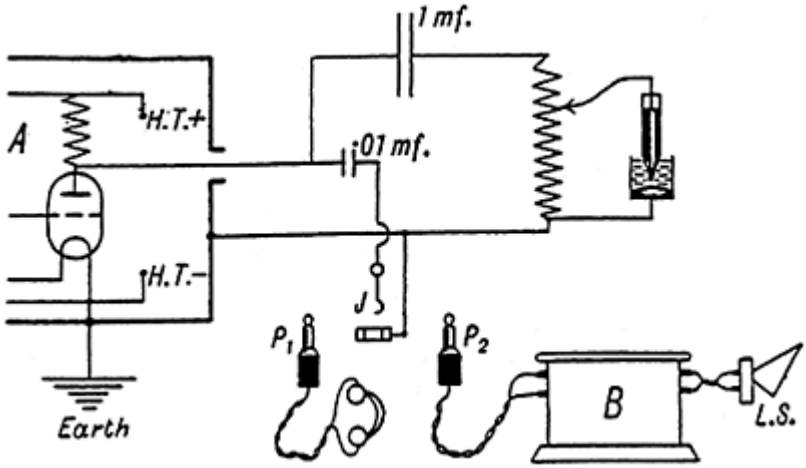


FIGURE A2.8 Amplifier and capillary electrometer used by Adrian to record nerve action potentials along with headphones, amplifier *(B)*, and loudspeaker *(LS)* used to aurally

monitor their frequency. (From Adrian and Bronk, 1928.)

> The three-valve amplifier owes much to the great kindness of Prof. Gasser, who supplied me with details of the amplifier used by him in America, and to the staff of Messers W.G. Pye and Co. of Cambridge, who redesigned an instrument on the same general line and planned the very compact and well-shielded lay-out of the apparatus.

Adrian provided details on his capillary electrometer by stating, “The capillary tube at present has a diameter of 0.03 mm at its working part, and a pressure of 140 mmHg is needed to bring the mercury to this point.” The working part refers to the sulfuric acid-mercury interface, the contour of which changes when current traverses it. Adrian stated that he could detect a voltage as small as 0.01 mV when his electrometer was connected to the three-stage amplifier.

Adrian knew that a photographic recording of the change in contour of the meniscus of the capillary electrometer was not a true representation of the applied voltage; this fact had been pointed out by Burch [1892]. Keith Lucas, a collaborator, devised a mechanical instrument for correcting capillary electrometer records. However, Adrian was less interested in waveform than in the presence or absence of action potentials.

Action Potentials in Afferent Fibers

In the paper that described the three-valve (vacuum-tube) amplifier and capillary electrometer, Adrian [1928] presented numerous examples of the nature of the action potentials in afferent nerve fibers in the frog sciatic nerve when the gastrocnemius muscle was stretched with known weights. He found that the frequency of the action potentials was related to the weight. Fig. A2.9 shows his corrected capillary electrometer records of this experiment.

After applying strict criteria to test the validity of his results, Adrian then recorded trains of afferent impulses in the saphenous nerve of a decapitated cat when a forcep was used to pinch the skin of the foot. The same result was obtained with a pin prick. He then recorded afferent impulses in the vagus nerve in the spinal and decerebrate cat and in the anesthetized rabbit. He also recorded trains of action potentials in the vagus nerve that were synchronous with the heartbeat and with respiration. Commenting on the latter he wrote: “The striking result is the absence of any sign of renewed discharge of impulses at the moment when the lungs are most deflated.”

In the summary to his paper, Adrian stated:

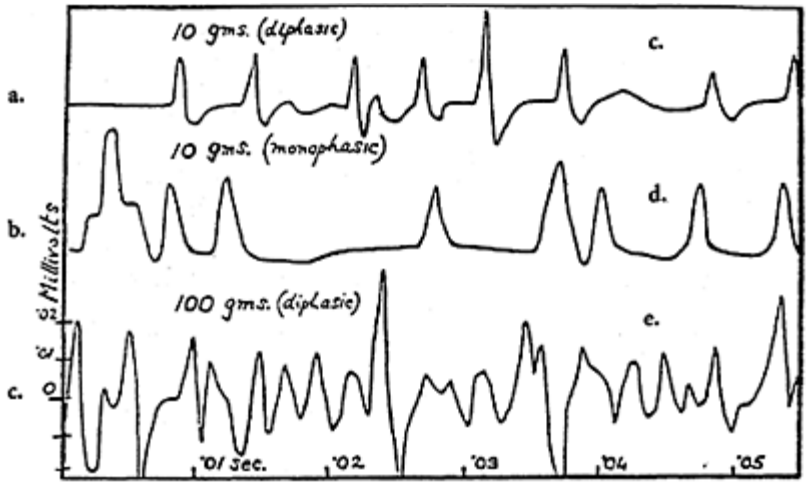


FIGURE A2.9 Corrected capillary electrometer records of afferent action potentials in a frog sciatic nerve when weights were applied to the gastrocnemius muscle, *(c)* shows the action potentials from 10 g applied for 10 s; *(d)* shows the same weight applied for 24 s; and *(e)* shows the action potentials for 100 g weight applied for 10 s. Note the higher frequency of action potentials with 100-g weight. (From Adrian, 1928.)

> It is probable that many of the oscillations represent action currents in a single nerve fibre, and these same general form and the same general time relations (allowing for temperature differences) in all sensory nerves in which they can be isolated sufficiently for measurement. There is no evidence that an increase in the stimulus increases the size of the action currents in single fibres, but the frequency of the impulses in the nerve trunk increases and leads to interference and overlapping of impulses in different fibres.

He continued:

> More detailed analysis of the results is postponed until experiments have been made on preparations containing a known number of sensory endings, if possibly only one.

It was not long before Adrian succeeded in recording action potentials in a single nerve fiber. Adrian and Bronk [1928] stated:

> To be sure of what is happening in the single nerve fibre, we have to devise a method which will put out of action all the other fibres in the nerve. The recording of impulses in a single fibre presents no difficulty, but the problem of isolating the fibre seemed much more formidable. Fortunately this has turned out not nearly so difficult as we had imagined.

Adrian and Bronk [1928] used the cervical rabbit phrenic nerve and dissected single fibers free with a needle and viewed the dissection with a binocular microscope. Teasing out a selected single fiber and sectioning it, electrodes were placed at the distal end. In this way, only afferent information was recorded. The equipment that they used in Fig. A2.8. Commenting on use of the headphones or a loudspeaker, they stated:

> The amplified action currents can be photographed with the capillary electrometer, but until the final stages are reached, it is usually more convenient to lead them to a telephone or loudspeaker and estimate the character of the discharged by the ear instead of the eye.

They continued by showing how aural monitoring aided in the experiment:

> When only a few fibres are in action, the electrometer excursions may be too small to detect on a screen, but they produce a series of faint clicks in the loudspeaker, and it is thus possible to control

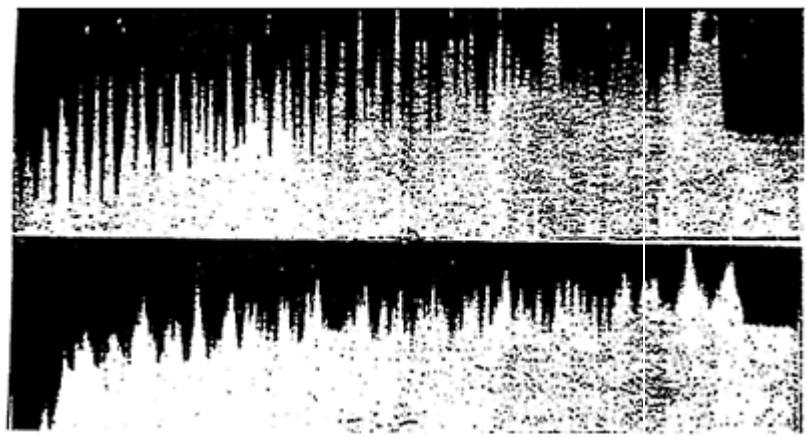


FIGURE A2.10 Capillary electrometer records from the phrenic nerve of the decerebrate cat breathing spontaneously, *(a)* The airway is open,

(b) The trachea was clamped. (From Adrian and Bronk, 1928.)

> the dissection, to expose a plate at the moment when the discharge is at its height, etc., without the inconvenience of wearing telephones.

Adrian and Bronk [1928] investigated the frequency of action potentials in an intact phrenic nerve of the decerebrate cat breathing spontaneously. The single-state amplifier (Fig. A2.8) was used. Figure A2.10*a* illustrates the phrenic nerve action potentials during spontaneous breathing, and Fig. A2.10*b* shows the recordings with the trachea clamped, the breathing becoming labored.

Having shown that the frequency of the action potentials in the phrenic nerve increased with depth of breathing, Adrian and Bronk [1928] set themselves the task of investigating how intrapulmonic pressure was related to the frequency of stimuli applied to the phrenic nerve. Using a rabbit in which the central circulation was occluded above C1, they connected a manometer to the trachea that could be clamped so that the manometer read the negative intrapulmonic pressure when the C4 phrenic nerve root was stimulated. They used a coreless induction coil connected to a rotary switch (rheotome) to enable delivery of stimuli at any desired frequency. By delivering bursts of stimuli of different frequencies, they plotted the negative intrapulmonic pressure as a function of frequency, clearly demonstrating the dependence of the amplitude of inspiration on the frequency of phrenic nerve stimuli.

Adrian published two monographs on his work; one was entitled, *The Mechanism of Nervous Action*, and the other, *The Basis of Sensation*. Interestingly, neither publication contains illustrations obtained with the cathode-ray oscilloscope; however, there are string galvanometer and capillary electrometer recordings. Obviously, Adrian had brought the capillary electrometer to a high degree of perfection, but it is important to remember that Adrian's interest lay in the frequency of the action potentials, not the intimate details of their waveforms.

Perhaps the best summary of Adrian's contributions appears in a single sentence in *The Mechanism of Nervous Action*. The section, entitled "Gradation of Activity," states:

> There is certainly no evidence to suggest that the impulses are graded in size, for the fact that sensory messages may produce a small or large effect according to the intensity of the stimulus is naturally explained by the varying number of fibres in action and by the varying frequency of the discharge in each fibre.

In 1932, Adrian shared the Nobel prize in physiology and medicine with Sherrington. The citation read, "For their discoveries regarding the function of the neuron."

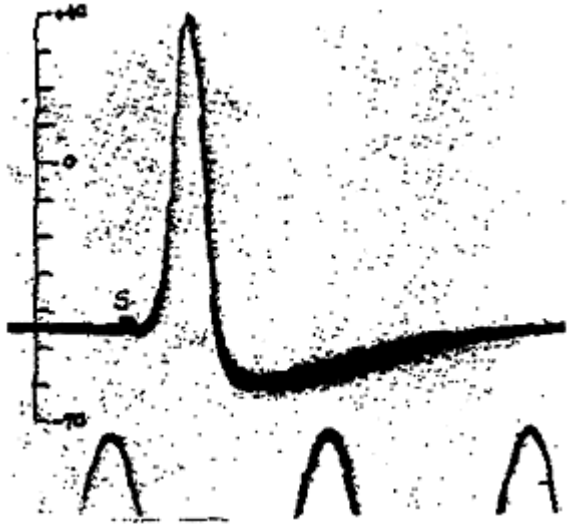


FIGURE A2.11 The transmembrane potential of the giant axon of the squid at rest *(left)* and during activity. Time marks 2 ms. (From Hodgkin and Huxley, 1939. With permission.)

Code of the Nervous System

From the impressive research performed by Gasser and Erlanger and by Adrian, the code of the nervous system was discovered. Stated simply, (1) intensity is signaled by the frequency of action potentials in a single axon, the action potentials all being the same, and (2) intensity is also signaled by the number of axons transmitting the information. In modern terms, the nervous system is a communications system that is binary and frequency modulated.

True Form of the Nerve Action Potential

Although the studies by Bernstein, Gasser, Erlanger, and Adrian provided information on the nerve action potential, its true form could not be established until the micropipet electrode was used to measure the transmembrane potential. Hodgkin and Huxley [1939] obtained the true waveform of the nerve action potential and thereby ushered in the modern era of electrophysiology. In a paper published in *Nature* on October 21, 1939, they reported that J.Z.Young [1930] called their attention to the giant axon (500 μm in diameter) of the squid, which was ideal for electrophysiologic studies.

The first micropipet electrode used by Hodgkin and Huxley consisted of a glass tube, 100 μm in diameter, that was slipped into the cut end of the giant axon. The electrode was mounted vertically and filled with seawater, and a silver-silver chloride electrode was inserted. The axon was then dipped into a container of seawater in which a second chlorided silver electrode was placed. The transmembrane potential was found to be −45 mV at 20°C. When the axon was stimulated, the action potential was 90 mV. Figure A2.11 is a reproduction of this historic record.

The Hodgkin and Huxley experiment clearly demonstrated two important phenomena: (1) there is a measurable resting transmembrane potential, and (2) the action potential is larger than the resting transmembrane potential, the latter indicating that activity represents more than a mere disappearance of the transmembrane potential. The same team later provided the explanation with their classic papers on ion fluxes. Today, the action of many drugs is explained on the basis of ion fluxes.

References

Adrian ED, Bronk DW. 1928. The discharge of impulses in motor nerve fibres: 1. Impulses in single fibres of the phrenic nerve. *J Physiol* 66:81.

Adrian ED. 1928. *The Basis of Sensation.* New York, WW Norton.

Bernstein J. 1868. Über den zeitlichen verlauf der negativen Schwankung des nervenstroms. *Arch Ges Physiol* 1:73–207.

Burch GJ. 18920. On the time relations of the excursions of the capillary electrometer. *Phil Trans R Soc* (Lond) 83A:81.

Burch GJ. 1892b. On a method of determining the value of rapid variations of potential by means of the capillary electrometer communicated by J.B.Sanderson. *Proc R Soc* (Lond) 48:89.

DeForest L. 1907, 1908. U.S. patent nos. 841, 387 (1907) and 879,532 (1908).

Einthoven W. 1903. Ein neues Galvanometer. *Ann Phys* 12(suppl 4): 1059.

Erlanger J, Gasser HS. 1924. The compound nature of the action current of nerve as disclosed by the cathode ray oscillograph. *Am J Physiol* 70:624.

Erlanger J, Gasser HS. 1937. *Electrical Signs of Nervous Activity*. Philadelphia, University of Pennsylvania Press.

Forbes A, Thacher C. 1920. Amplification of action currents with the electron tube in recording with string galvanometer. *Am J Physiol* 56:409.

Gasser HS, Erlanger J. 1922. A study of the action currents of nerve with the cathode ray oscillograph. *Am J Physiol* 62:496.

Gasser HS, Newcomer HS. 1921. Physiological action currents in the phrenic nerve. *Am J Physiol* 57(1):1.

Geddes LA, Foster KS, Senior J, Kahfeld A. 1989. The inductorium: The stimulator associated with discovery. *Med lustrum* 23(4):308.

Gotch F, Horsley V. 1888. Observations upon the electromotive changes in the mammalian spinal cord following electrical excitation of the cortex cerebri. *Proc R Soc* (Lond) 45:18.

Helmhotz H. 1850. Note sur la vitesse de propagation de 1'agent nerveux dans les nerfs rachidiens. *C R Acad Sci* 30:204.

Helmholtz H. 1951. Note sur la vitesse de propagation de l' agent nerveux. *C R Acad Sci* 32:262.

Helmholtz H. 1853. On the methods of measuring very small portions of time and their application to physiological purposes. *Phil Mag J Sci* 6:313.

Hodgkin AL, Huxley AF. 1939. Action potentials recorded from inside a nerve fiber. *Nature* 144:710.

Hoff HE, Geddes LA. 1957. The rheotome and its prehistory: A study in the historical interrelation of electrophysiology and electromechanics. *Bull Hist Med* 31(3):212.

Hoff HE, Geddes LA. 1960. Ballistics and the instrumentation of physiology: The velocity of the projectile and of the nerve impulse. *J Hist Med All Sci* 15(2):133.
Young JZ. 1930. Structure of nerve fibers and synapses in some invertebrates. *Cold Spring Harbor Symp Quant Biol* 4:1.

Index

G

O