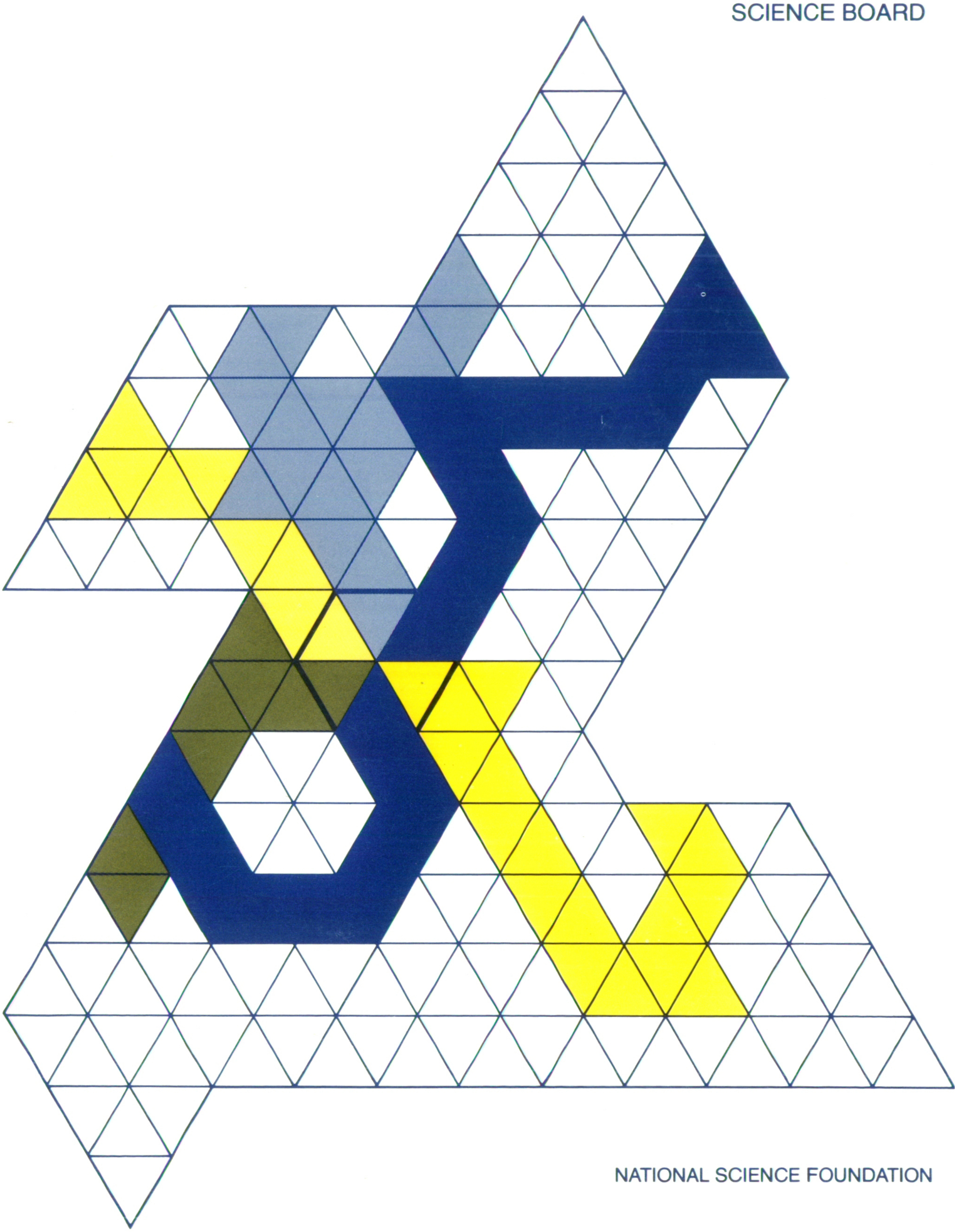
ONLY ONE SCIENCE

TWELFTH
ANNUAL REPORT
OF THE NATIONAL
SCIENCE BOARD



NATIONAL SCIENCE FOUNDATION

ONLY ONE SCIENCE TWELFTH ANNUAL REPORT OF THE NATIONAL SCIENCE BOARD

MEMBERS OF THE NATIONAL SCIENCE BOARD

DR. LEWIS M. BRANSCOMB (Chairman, National Science Board), Vice President and Chief Scientist, International Business Machines, Inc., Armonk, New York

MR. HERBERT D. DOAN (Vice Chairman, National Science Board), Chairman, Doan Resources Corporation, Midland, Michigan

DR. RAYMOND L. BISPLINGHOFF, Vice President for Research and Development, Tyco Laboratories, Inc., Exeter, New Hampshire

DR. LLOYD M. COOKE (Vice Chairman, Committee on Twelfth NSB Report), President, National Action Council for Minorities in Engineering, Inc., New York, New York

DR. EUGENE H. COTA-ROBLES, Professor of Biology, Biology Board of Studies, University of California at Santa Cruz, Santa Cruz, California

DR. PETER T. FLAWN, President, University of Texas at Austin, Drawer T, Austin, Texas

DR. ERNESTINE FRIEDL, Dean of Arts and Sciences and Trinity College, Professor of Anthropology, Duke University, Durham, North Carolina

DR. MARY L. GOOD, Vice President, Director of Research, UOP, Inc., Des Plaines, Illinois

DR. JOHN R. HOGNESS (Chairman, Committee on Twelfth NSB Report), President, Association of Academic Health Centers, Washngton, D.C.

DR. WILLIAM F. HUEG, JR., Professor of Agronomy and Deputy Vice President and Dean, Institute of Agriculture, Forestry, and Home Economics, University of Minnesota, St. Paul, Minnesota

DR. MICHAEL KASHA, Distinguished Professor of Physical Chemistry, Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida

DR. MARIAN E. KOSHLAND, Professor of Bacteriology and Immunology, University of California at Berkeley, Berkeley, California

DR. PETER D. LAX, Professor of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, New York

DR. WALTER E. MASSEY, Director, Argonne National Laboratory, Argonne, Illinois

DR. HOMER A. NEAL, Provost, State University of New York at Stonybrook, New York

DR. MARY JANE OSBORN, Professor and Head, Department of Microbiology, University of Connecticut School of Medicine, Farmington, Connecticut

DR. JOSEPH M. PETTIT, President, Georgia Institute of Technology, Atlanta, Georgia

DR. DAVID V. RAGONE, President, Case Western Reserve University, Cleveland, Ohio

DR. DONALD B. RICE, JR., President, The Rand Corporation, Santa Monica, California

DR. STUART A. RICE, Frank P. Hixon Distinguished Service Professor of Chemistry, The James Franck Institute, University of Chicago, Chicago, Illinois

DR. ALEXANDER RICH, Sedgwick Professor of Biophysics, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

DR. EDWIN E. SALPETER, J. G. White Professor of Physical Sciences, Cornell University, Ithaca, New York

DR. CHARLES P. SLICHTER, Professor of Physics and in the Center for Advanced Study, University of Illinois, Urbana, Illinois

Ex Officio

DR. JOHN B. SLAUGHTER, Director, National Science Foundation, Washington, D.C.

Executive Secretary

MISS VERNICE ANDERSON, National Science Board, Washington, D.C.

IMMEDIATE PAST MEMBERS OF THE NATIONAL SCIENCE BOARD

DR. RICHARD C. ATKINSON (Past Member *ex officio*), Chancellor, University of California at San Diego, La Jolla, California

DR. JEWEL PLUMMER COBB, President, California State University at Fullerton, Fullerton, California

DR. NORMAN HACKERMAN (Past Chairman, National Science Board), President, Rice University, Houston, Texas

DR. W. N. HUBBARD, JR., President, The Upjohn Company, Kalamazoo, Michigan

DR. SAUNDERS MAC LANE, Max Mason Distinguished Service Professor of Mathematics, Department of Mathematics, University of Chicago, Illinois

DR. GROVER E. MURRAY (Past Vice Chairman, National Science Board), University Professor, Texas Tech University and Texas Tech University School of Medicine, Lubbock, Texas

DR. L. DONALD SHIELDS, President, Southern Methodist University, Dallas, Texas

DR. JAMES H. ZUMBERGE, President, University of Southern California, Los Angeles, California

ONLY ONE SCIENCE

TWELFTH ANNUAL REPORT OF THE NATIONAL SCIENCE BOARD

To him who devotes his life to science,
nothing can give more happiness
than increasing the number of discoveries.
But his cup of joy is full
when the results of his studies
immediately find practical application.

There are not two sciences.
There is ONLY ONE SCIENCE
and the application of science,
and these two activities are linked
as the fruit is to the tree.

Louis Pasteur

NATIONAL SCIENCE FOUNDATION

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Letter of Transmittal

September 28, 1981

My Dear Mr. President:

I have the honor of transmitting to you and through you to the Congress the Twelfth Annual Report of the National Science Board—*Only One Science.*

As the title implies, this Report is a departure from previous ones. In a narrative form it tells about scientific research and describes how the results of such research affect and benefit society. The Report does not make specific policy or budget recommendations.

The introduction to the Report quotes Louis Pasteur: "There is *Only One Science* and the application of science, and these two activities are linked as the fruit is to the tree." The Board believes that this is true. We also believe that it is of great importance for all Americans to appreciate how research, technological development, and human welfare are inevitably and necessarily interrelated and intertwined. These stories should help in achieving that understanding.

In preparing the Report, the Board selected as subjects six fields of scientific endeavors from a long list of potential choices. Some, if not all, of these stories should appeal to readers with different interests. All of the chapters—whose subjects range from how the seismic system is used in exploring for gas and oil to the uses of X rays in medical diagnosis—are of current interest.

The enabling legislation which created the National Science Foundation mandates the Foundation "To promote the progress of science; to advance the national health, prosperity, and welfare, to secure the national defense and other purposes." In an informal way this Report deals with the relationship of science to the "general welfare." We hope the stories are interesting—the record of science stands for itself.

Respectfully yours,

Lewis M. Branscomb

Lewis Branscomb
Chairman, National Science Board

The Honorable
The President of the United States

Preface

Each year the National Science Board submits to the President and through him to the Congress a report which deals with issues of concern to the Board and to the National Science Foundation. Every other year the report is entitled *Science Indicators;* it consists of an updated series of data and indices which portray the status of science and technology in the United States. In intervening years, reports of the National Science Board have dealt with a wide range of scientific issues which the Board feels are important.

The Twelfth Board Report represents a significant departure from previous reports in that it attempts, in a narrative fashion, to deal with a somewhat broader aspect of the National Science Foundation's responsibilities. In establishing the Foundation in 1950, the Congress, among other things, directed the National Science Foundation and the Board to *"appraise the impact of research* upon industrial development and *upon the general welfare."* This Report was developed to do that through the use of historical, anecdotal stories of discoveries in six different representative fields of science.

Although the subject of each chapter is different, the story each tells is characterized by change. To look at any of the fields at a given moment in the past and to try to project how a thread of ideas, activities, and circumstances would unfold in the future, would be difficult, if not impossible. However, the perspectives gained from the Report should give readers a keener sense of the interactions among scientific research, industry, academia, and the individual. The Report should not be viewed as a comprehensive statement of the current status of a particular field or as a basis for future public policy—except to the general extent that "past is prologue" to the future.

In the course of preparation of the Report, various aspects of it were reviewed by the current Members of the National Science Board and by the immediate past Board Members, all of whom are listed inside the front cover of this Report. The Committee responsible for the preparation for the Twelfth Board Report wishes to acknowledge the extensive participation of all present and immediate past Members of the Board and wishes particularly to express appreciation to Dr. Gwynn C. Akin, Staff Director for the Report and Consultant to the National Science Board, who was the principal staff architect of the Report. The Committee also wishes to express thanks to Dr. Carlos Kruytbosch, the Executive Secretary to the Committee and to Dr. Allen Shinn who served in that same capacity in the early months. In addition, more than 250 individuals provided information, documentation, written text, advice, suggestions, critical comments, or technical assistance during the course of preparation of the Report. Their names are listed in the Acknowledgment section in the back of this volume. If any one has been omitted, we offer our sincere apologies.

COMMITTEE ON THE TWELFTH BOARD REPORT

John R. Hogness, Chairman
Lloyd M. Cooke, Vice Chairman
Herbert D. Doan
Walter E. Massey
Joseph M. Pettit

Contents

Introduction

To him who devotes his life to science, nothing can give more happiness than increasing the number of discoveries. But his cup of joy is full when the results of his studies immediately find practical application.

There are not two sciences. There is ONLY ONE SCIENCE *and the application of science, and these two activities are linked as the fruit is to the tree.*

Louis Pasteur

No age in history has come closer to providing Pasteur's link than the twentieth century. In a world of explosive changes, the extraordinary growth of science and technology has had an impact on everyone and on every aspect of life. The results of scientific advances—automobiles, airplanes, wash-and-wear clothes, antibiotics, television, glass windowpanes, air conditioners—are pervasive and ubiquitous. And yet they are often taken for granted or go completely unnoticed.

Dealing with the rapid rate of scientific and technological change and with its results presents a unique challenge—a challenge which can be met only by strengthening the bridge of understanding between the individual and the world of science. The public understanding of the purposes and effects of science and technology is essential to the health and vitality of a modern society.

This realization has led the National Science Board to take a novel approach in its Twelfth Report. Rather than presenting a formal guide to policy, the Report describes, in informal, narrative style, how certain scientific discoveries occurred and how they have affected society. The Report examines six topics of importance in which research, technological innovation, public need, and human welfare have, over varying periods of time, come together to create a body of scientific knowledge and a related technology that enhance the quality of life.

The Report is meant to show that, in the words of Jules Henri Poincaré, "Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house." Even the failures that are inevitable in research provide new insights. The contamination of a culture plate resulted in the finding of the antibacterial properties of penicillin; a disproven hypothesis can lead to a fresh view of a problem under study.

The six topics examined (computers and semiconductors, pesticides and pest control, the seismic system, survey research and opinion polls, synthetic fibers, and X rays for medical diagnosis) represent a spectrum of scientific endeavors in the physical, biological, medical, and social sciences. The topics were selected from a group which met several criteria: a significant impact on society and the individual, a substantial information base, a history sufficient to assess their development, a firm relationship to scientific research, and general current interest. Each chapter illustrates the problems of dealing with the technological, social, and political realities of society and demonstrates the strong recurring links among accomplishments by individuals, universities, industries, and the government that bring about the introduction to society of almost every major scientific development.

One striking theme that recurs throughout these chapters is that approaches to solutions of scientific problems are diverse and varied. The popular concept of scientific and technological development is that of a "clean" linear progression from basic, "nontargeted" research leading to applied research which, in turn, leads to technological development and the marketing of a product. This orderly progression occurs only rarely. When this direct linear relationship between a basic observation and the practical application does exist, the connection may be made within a matter of months or years—or the fundamental observation may lie dormant for centuries before its uses can be perceived and a new technology developed. Alternatively, a breakthrough in technology may stimulate basic research, which in turn allows the investigator to delve deeper into nature's secrets, thus generating more knowledge that can be applied to improve the technology. Or a need for a particular technology may be perceived, although the basic knowledge necessary for its development is not yet available. The need for the technology may stimulate fundamental research which permits the technological development which may, in turn, provoke additional basic research.

Although basic research often leads to technological application and technological application fosters basic research, serendipity is also an ingredient of research. It is the "X" factor that is both unpredictable and often integral to discovery. It is the chance observation that stimulates a new idea or a new scientific application. But here, too, new ideas will occur as a result of such a circumstance only when a background of knowledge is already present. For, as Pasteur observed, "In the fields of observation, chance favors only the mind that is prepared."

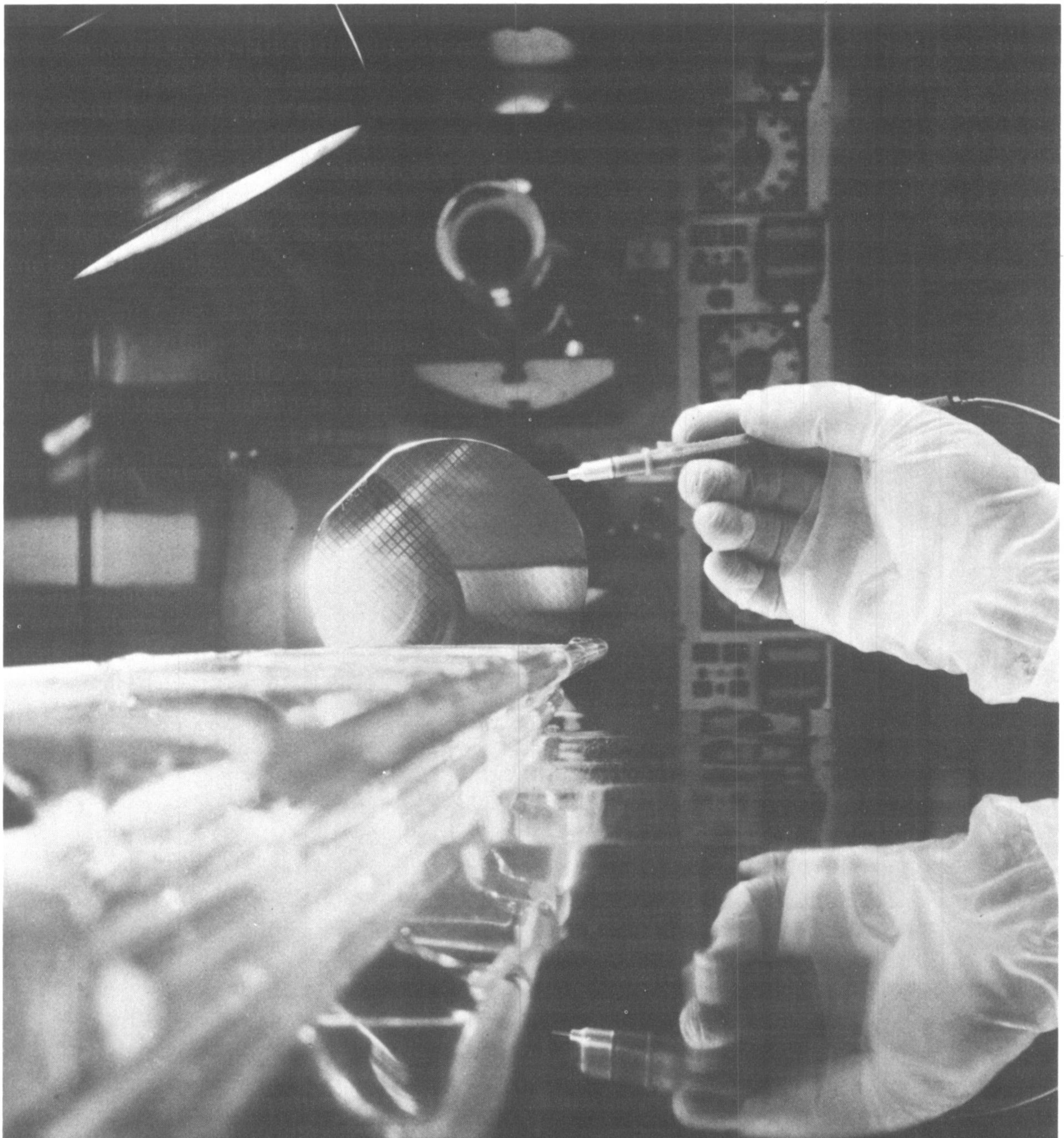
One of the early examples of the use of seismic waves came from a military officer's trying to discover if the enemy were digging a secret tunnel. The officer's basic tools were a pan of water, set flat on the ground, and a sharp-eyed soldier stationed to look for ripples in the water caused by digging. Today, the interlocking combinations of basic research, chance observation, need, and technological development have replaced that soldier with a host of trained geophysicists and that pan of water with sophisticated recording instruments. About the only thing that has not changed since that soldier's lonely vigil is a strong urge, sometimes even a need, to know what is going on under the earth's surface. Whether it is the search for new energy resources or for a better way to predict earthquakes, additional basic research and new technology keep driving each other to fulfill an age-old quest for answers.

Recently, from data developed using the seismograph and other technologies, a new scientific theory was born—plate tectonics. According to this concept the crust of the earth is comprised of about 20 plates that are in continual motion. (Plates are huge blocks of the earth's crust that float on the dense, hotter, more fluid rock below.) The knowledge gained from studying this new field has, in turn, provided a greatly improved understanding of the movement of the continents, volcanic activity, and earthquakes. Perhaps this new concept will not provide much practical application for many years. But then it is unlikely that the person who conceived the first abacus more than 2,000 years ago could have dreamed of a pocket-size calculator complete with 30 mathematical functions and a memory core.

It was not until this Report was well underway that members of the Board became aware of another type of interesting interrelationship among the topics. There are many cross-connections between the developments in the various areas of study; a technology discussed in one of the chapters often has significant applications to the technologies discussed in others. For example, although the discussion of X rays deals mainly with medical diagnosis, radiation is also used in pest control to sterilize male insects or to develop mutations in seeds so that pest-resistant crops can be grown. A clear relationship also exists between the seismic system and synthetic fibers—oil and gas discovered by using that system provide the base substances for the production of synthetic fibers. The most impressive interconnection, however, is the relationship of the computer to the other subjects discussed. For example, computers are an integral part of the development of the CT (computerized tomography) scanner so important in radiological diagnosis today. Many advances in the use of the seismic system to help locate oil and gas could not have occurred without computers. Computers are used in analyzing data from large scale statistical surveys. They are even used in the control of production processes in the manufacture of synthetic fibers and in numerous calculations necessary for the success of certain pest control techniques. One technological development often makes another development possible, and so science builds upon science, and one technology on another.

Discovery is a product of opportunity, imagination, brilliance, persistence, and serendipity. Unfortunately, none of these factors can be measured exactly, nor can they be ordered at will. An enlightened society must recognize the need for major opportunities in unfettered research in which imaginative scientists feel free to pursue their curiosity beyond the limits of current knowledge.

To extend Pasteur's analogy of the fruit and the tree, if society expects a bountiful harvest, it must constantly nurture and feed the "tree" of science. It must also remember that no one can predict exactly what or when this particular tree will bear—only that, in time, it will.



Computers and Semiconductors

A few summers ago, the 22-year-old victim of an automobile accident was rushed to Methodist Hospital of Indiana with critical head injuries. During the next 2½ days, while the patient's respiratory and circulatory systems were maintained by machines, doctors ran a variety of sophisticated tests but failed to find even a glimmer of neurological activity. The patient remained in a coma. Finally, doctors concluded that he had suffered total brain death.

Because the victim had been in otherwise excellent health, the doctors brought up the possibility of using his kidneys for transplant purposes. The next-of-kin agreed to the donation—the process of finding suitable recipients began.

Doctors had to find two people who were in critical need of a kidney transplant and whose immune systems were compatible with that of the donor. If the blood and tissue antigens of the donor and receiver are not the same or very similar—and there are tens-of-thousands of possible combinations—chances are overwhelming that the recipient's immuno-defense mechanism will attack, and probably reject, the donor's kidney.

In an effort to find a sufficiently close match, the donor's types were keyed into a terminal at the hospital and transmitted to a computer operated by the South-Eastern Organ Procurement Foundation (SEOPF) in Richmond, Virginia. There, a 10-second search of almost 1,500 possible candidates registered by 40 transplant centers in a 17-state area produced a printout of more than 100 names.

silicon wafer (in the fore-
und) is being loaded into a
y with a vacuum pencil. A
fusion furnace used in inte-
ted circuit fabrication is in
background.
ional Semiconductor Inc.

Although all the people on the list had the same blood type and one or more matching tissue antigens, only two of them met enough transplant criteria to be considered as prime prospects: a 34-year-old Indianapolis housewife who had been receiving dialysis treatments for the past 7 years and a 40-year-old businessman in Pittsburgh, Pennsylvania, who was failing on his dialysis treatments and was in desperate need of a transplant.

As soon as the SEOPF computer relayed this information to Indianapolis and to the transplant centers where the recipients were registered, a surgical team removed the accident victim's kidneys, flushed them with a preserving solution, and packed them in ice-filled containers. One was raced across town in an ambulance; the other was hand-carried to Pittsburgh on a commercial jet. Less than 8 hours after the kidneys had been removed from the victim, a message was flashed by the computer to all terminals in the SEOPF network: *Two kidneys transplanted with successful results.*

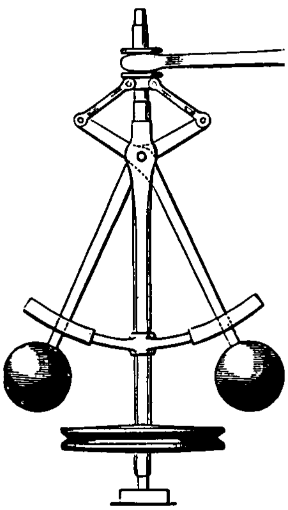
While transplant-kidney-matching is one of the more peripheral computer applications, it does illustrate how far the state of the art has come in just 30 years. The early vacuum-tube computers of the late 1940s and early 1950s were monstrous machines that were clearly the exclusive province of a handful of mathematicians, physicists, engineers, and astronomers. Even the next two generations of data processing machines, built and marketed during the 1960s, could be used efficiently only by people with a strong background in mathematics and science. Today, however, more and more computer capability is found in the hands of people whose prime interests are in other areas. Computers are used routinely in medicine, in schools, in factories, in offices, and in the home. And while these machines have not yet become everyone's tool, there is no doubt that the current generation of computers—compact, solid-state systems relying heavily on microtechnology—is far faster, more versatile, less expensive, and easier to use than its predecessors.

THE ORIGINS OF DATA PROCESSING

Many people outside the electronics field tend to think of the computer as a magical engine which sprang into existence full-blown, like Athena from the head of Zeus. In fact, the computer has roots that reach far back in history—more than 2,000 years—to the abacus. Computer technology rests upon knowledge accumulated over many centuries and evolved from the genius of many inventive minds. No single breakthrough or classic experiment brought this device into being; instead, historians trace the development of computers along at least four technical tributaries which merged about 40 years ago to form one powerful stream.

Automatic Regulators

One tributary, often overlooked in recounting the background of computers, goes back to James Watt and his steam engine. In 1788 when Watt applied a centrifugal fly-ball governor to his engine, he did much more than just improve the machine's efficiency. In effect, he showed for the first time how a machine could "examine" its own output and use the information to monitor and control its internal operation.



In 1788 James Watt designed the fly-ball governor to regulate the speed of his steam engine and to allow it to function independently of load changes.
Smithsonian

Other automatic regulators followed. In 1830 Andrew Ure invented the thermostat to help control the temperature of furnaces; in 1852 Leon Foucault devised the gyroscope, first used to maintain the course of torpedoes, but later to become the mainstay of navigation for ships and airplanes. In the 1860s James Clerk Maxwell supplied a mathematical theory which, among other things, helped to establish the science of automatic controls.

Today there are feedback mechanisms on everything from dishwashers to data processing machines. In concept, however, they all go back to Watt's fly-ball governor, a system of revolving weights that act as an automatic throttle. Interestingly, Watt was not searching for an abstract principle when he built the control mechanism. Rather, he was looking for a practical answer to an urgent need—a device that would enable his steam engine to maintain operating speed despite changes in load.

Boolean Logic

Another stream which contributed to modern-day data processing techniques flowed from quite a different direction. Obviously, a computer does not depend on a fly-ball governor, thermostat, or gyroscope to self-direct its activities. But it does rely on something equally ingenious—a set of instructions, or a program, which governs the path of electronic signals through the machine's switching circuitry. And this ability to process a sequence of logical statements goes back to the work of a remarkable nineteenth-century figure named George Boole.

In 1854 Boole published *An Investigation of the Laws of Thought* which did for logic what Euclid had done for geometry. The book described how logical statements could be translated into precise mathematical forms. Boole's mathematical logic was binary in nature because it was based on the premise that statements are either true or false. The switching circuits of digital computers are also binary in nature because they can exist in only one of two states: "on" or "off."

What Boole did for the development of computers was to set the stage for the stored-program computer. In the earliest data processing machines, circuit patterns were more or less fixed to perform a specific job. The only way a program could be changed, for example, was literally to go in and reset the wires. In 1938, however, mathematician Claude Shannon of the Massachusetts Institute of Technology (MIT) suggested in a technical paper that Boolean algebra could be used for more flexible programming of electronic calculators. Eight years after that, the famed mathematician John von Neumann actually showed how programs could be written and stored in the machine.

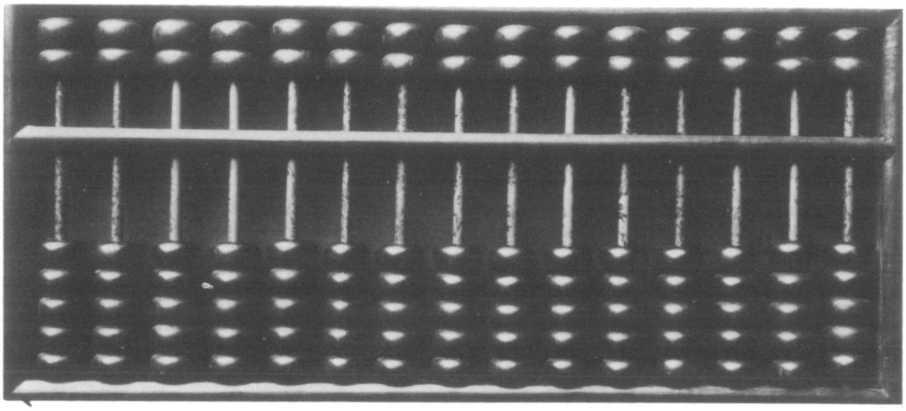
"This resulted in two very great gains," explains Christopher Evans in his book, *The Micro Millennium*.

> In the first place, one could take advantage of the computer's huge processing speed and allow it to change programs when required; it could switch from one program to another in a fraction of a second instead of relying on the lumbering skills of its attendant human being. In the second place, and this is far and away the most important point, it meant that programs within the system could interlock and interact.... In principle, programs could even modify other programs, rewriting them to fit the needs of the moment and integrating them with yet others within the suite.

Thus, in one conceptual jump, the feedback concept that began with Watt was joined with the mathematical logic of Boole.

Calculating Machines

A more direct tributary—one that has been recounted often in the history of data processing—sprang from the abacus. This little device came into being more than 2,000 years ago and still is the most widely used calculating tool on earth. In vast areas of Asia, in fact, it is virtually the only known counting machine.



The abacus is a device for making arithmetical calculations. Similar counting tools have been used throughout Asia and the Middle East for more than 2,000 years.

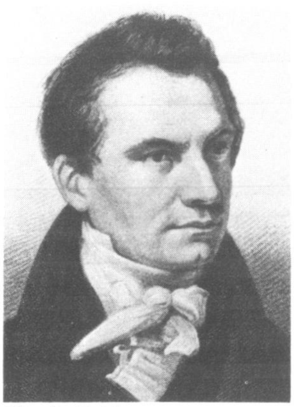
Smithsonian

The abacus was developed as a direct result of early efforts to count. Probably the first quantitative symbols used were a "two" and a "five": "two" because people have two hands and "five" because of the five digits on each hand. It is not surprising that the most popular form of abacus makes use of the two-five or binquinary notation system. The Chinese *suan-pan*, for example, consists of a series of rods on which beads are strung. There are seven beads on each rod—separated by a divider strip into five on one side (each bead equivalent to 1) and two on the other (each bead equivalent to 5). Thus, the number "seven" can be expressed on a line by moving over one bead from the "five" side and two beads from the "one" side.

In the hands of a skilled practitioner, the *suan-pan* is an amazingly fast, accurate, and versatile device. Nevertheless, it has a distinct shortcoming. It cannot carry over tens from one line to another, and as mathematical horizons were expanded, this deficiency became a major problem.

Although many counting machines were devised over the years, it was not until the seventeenth century that the next clear-cut advance came along. In 1642 a young Frenchman named Blaise Pascal, working in his father's tax office in Rouen, built an ornate shoebox-size device which employed linked gears to add, subtract, and, most importantly, carry over tens. Thirty years later Gottfried Leibniz improved on Pascal's machine so it could multiply, divide, and calculate square roots. In effect, these two inventors brought about the calculating age. (Pascal's contribution to survey research is mentioned in the chapter, "Survey Research and Opinion Polls.")

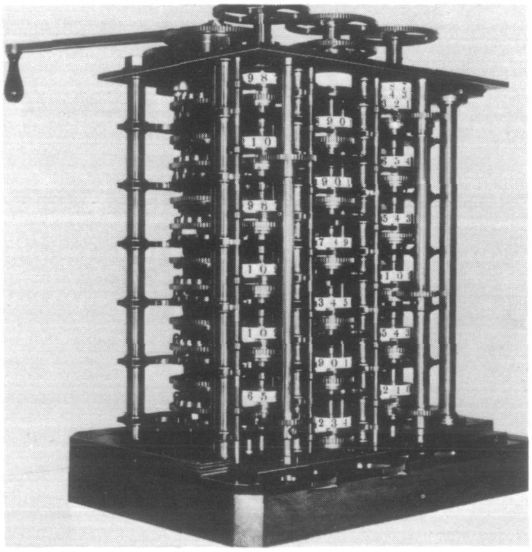
During the next one-half century many new calculating machines were built, but all were primarily modifications or refinements of Pascal's original design. In fact, although the next important development had nothing to do with calculating, it was vital to advancement of the science. In 1780 Joseph Marie Jacquard built an automatic weaving loom which operated from instructions punched into cards or paper tape. This invention, which revolutionized the weaving industry, led directly into one of the most unusual stories in the history of computers.



Charles Babbage
(1791-1871)
IBM Archives

The year was 1822 and the principal was a complex young Englishman named Charles Babbage. Babbage, a mathematician and inventor of the railroad cowcatcher and the first tachometer, was becoming increasingly incensed by the many errors he found in insurance records, logarithm tables, and other lists of data. His fetish for accuracy was so great, in fact, that after reading Lord Tennyson's famed line, "Every moment dies a man/Every moment one is born," he wrote to the poet, "It must be manifest that if this were true, the population of the world would be at a standstill." Babbage's recommended change to Tennyson, "Every moment dies a man/Every moment $1\frac{1}{16}$ is born."

In 1822 Babbage began work on a machine, which he called the "Difference Engine," that could help solve polynomial equations to six places. The English government was so impressed by the machine's potential for compiling navigational and artillery tables that they subsidized him heavily. The projected machine, said Babbage, would be able to do complex calculations and print out its results. It was to have a "memory" section made up of the same sort of punched cards used by Jacquard's loom; cards were also to have been used for input to the machine and control of its successive operations. The device was to have an arithmetic unit, called a mill, in which to store data; it was to be able to set up its own results in type, thus avoiding transcription errors.



This replica of Babbage's Difference Engine was constructed according to his plans. Unlike a true computer Babbage's invention was capable of making complicated computations, but it had no memory as a true computer does.

IBM Archives

Babbage was literally 130 years ahead of himself. He built a small working model, but never was able to complete a full-sized Difference Engine. The reason for his failure, however, had nothing to do with the concept; the machine could not be built primarily because the technology of the time was not adequate to permit construction of the needed parts. By the time the eccentric genius died in 1871, he had managed to put together only a few parts. Nevertheless, his elaborate drawings of the machine left no doubt that he was well on his way to a true computer.

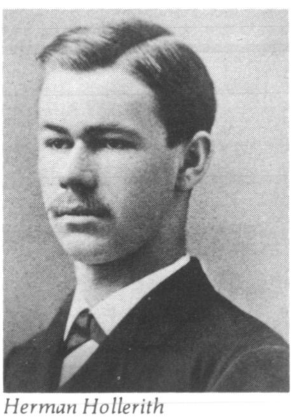
Several years ago, B. V. Bowden, writing in *Think* magazine, described his efforts to track down the Babbage story. Aside from the discovery of many papers which proved the genius of the farsighted Englishman, the trail led to Lady Lovelace who was the daughter of the famed poet Lord Byron. It seems that Lady Lovelace was as much a prodigy in mathematics as her father was a master of poetry. She devised, among other things, a form of binary arithmetic and an "infallible" system for predicting horse race winners. The binary system, originally described by Boole, has stood the test of time and is today used in electronic digital computers. The betting system cleaned out the family fortune and forced Lady Lovelace to pawn all her jewels.

Somewhere along the line Lady Lovelace met Charles Babbage. Apparently their mutual interest in mathematics was the spark for a long-lasting relationship. As Bowden wrote:

> Lady Lovelace often visited Babbage while he was making his machines, and he would explain to her how they were constructed and used. As one of her contemporaries recalled, "While the rest of the party gazed at this beautiful instrument with the same sort of expression that some savages are said to have shown on first seeing a looking glass or hearing a gun, [she] understood its working and saw the great beauty of the invention." She worked out some very complicated programs and would have been able to use any of the modern machines. She wrote sketches for several papers, but published only her notes on Babbage, and they were anonymous.

Perhaps the most perceptive observation made by Lady Lovelace was one which could just as easily be applied to the present generation of powerful microminiature computers. It concerned the question of whether Babbage's machine could be considered "creative." She wrote, "The Difference Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths. Its province is to assist us in making available what we are already acquainted with."

Electric Tabulating Machines



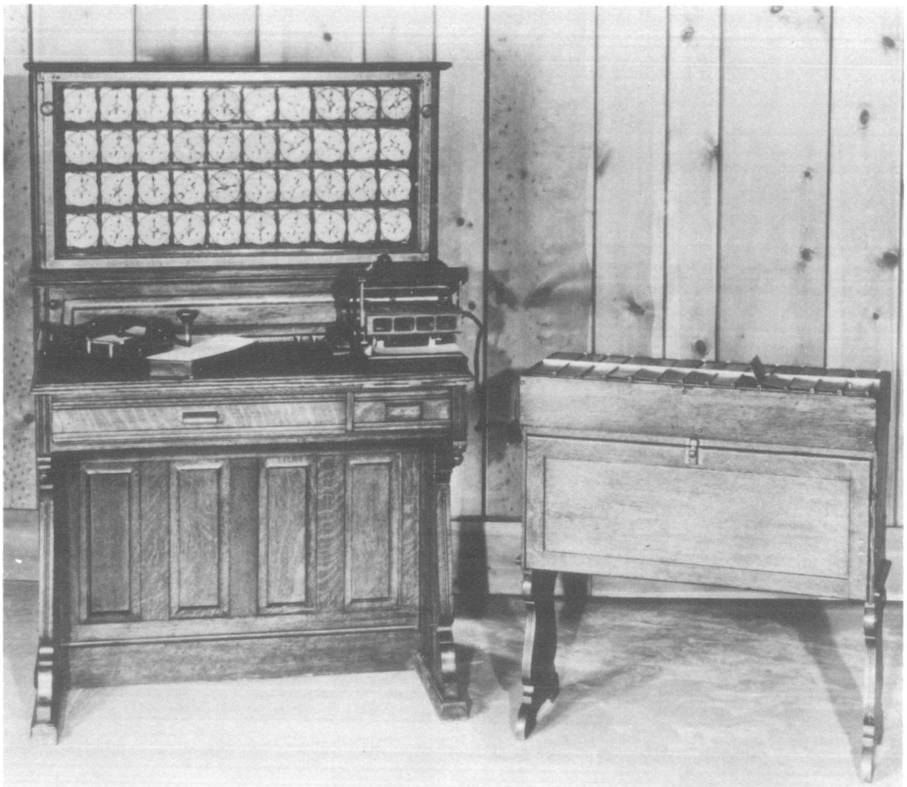
Herman Hollerith (1860-1929)

IBM Archives

Babbage, Boole, Lady Lovelace, Jacquard, Pascal, and many others brought technology to the brink of modern data processing. But it took the work of one man to push it over into the practical devices that are found today throughout the world. He was Herman Hollerith, a statistician from Buffalo, New York, who arrived on the scene just as the United States government was about to stagger its way through the once-a-decade census count.

The basic problem was simple: the government had needed 7 full years of counting and tabulating to complete the 1880 census. All data were handwritten on cards; the cards were manually sorted into various categories; each category was manually counted; and then the cards were resorted into new piles and counted once again. Compiling the census was such a large, tedious job that by the time the final count was completed, it was already outdated. With the time for the 1890 census count approaching and immigration swelling population ranks by the day, the Census Bureau could envision 9 or even 10 full years to take the next count.

Although only 40 years had elapsed between Babbage's frustrating efforts to build a Difference Engine and the development of the Census Bureau's forebodings over the 1890 census, two things had happened in the meantime to change the technological climate. First, both machining and manufacturing skills had improved tremendously as a result of the Industrial Revolution; and, second, an exciting new form of power, electricity, was now being used to drive an increasing number of machines.



Herman Hollerith's tabulator which he developed for use in the United States Census of 1890.
Smithsonian

Hollerith worked out an electromechanical method of recording, tabulating, and organizing census data. His system used cards, about the size of the old United States dollar bill, in which data were recorded in the form of holes made with a conductor's hand punch. The punched cards were then positioned one-by-one over mercury-filled cups in a special machine. At the touch of a lever, rows of telescoping pins descended on the cards; where there was a hole the pin simply dropped through into the mercury, thus completing an electrical circuit. The electrical impulse, in turn, was used to move a pointer one position on a dial. As the dials went around, various totals were accumulated. (Hollerith's work was also a key contribution to the field of survey research and is referred to in the chapter on that subject.)

While this method sounds primitive by today's standards, it enabled the government to complete the 1890 census of 62 million people in just one-third the time of the 1880 census of 50 million. When news of the system reached industry, an almost immediate demand developed. Before 1900 Hollerith was marketing his unit record machines to many of the nation's largest firms. The New York Central Railroad used them for car accounting; the Marshall Field Department Store installed electromechanical tabulators for sales analysis work; the Penn Steel Company in Philadelphia used them for cost accounting; and the Western Electric Corporation installed several of the machines for sales analysis. Meanwhile, Hollerith went to Czarist Russia to set up a similar system for that country's census count.

Adoption of Hollerith's machines took place at a time when the United States was embarking on an unprecedented technological surge. From about 1880 to 1930 commerce in America expanded at a prodigious rate. The railroads pushed west, north, and south to open new markets and create new industries. Manufacturing firms adopted mass production techniques which helped to increase their productivity many times over. The concept of interchangeable parts ensured that a product manufactured or purchased in one part of the country could be serviced in another. American ingenuity produced more than 1.3 million new patents in the first third of the twentieth century alone.

In many respects, this was an ideal environment for the introduction of data processing techniques. Every bit of this industrial and commercial development created accounting and recordkeeping problems. In 1911 Hollerith merged his young company with two other firms to become the Computing-Tabulating-Recording (C.T.R.) Company. C.T.R. (which eventually became the International Business Machines Corporation) had four basic units to offer: a key punch for putting holes in cards; a hand-operated gang punch for coding repetitive data into several cards at the same time; a vertical sorter for arranging cards in selected groups; and a tabulating machine for compiling the data punched into cards.

Over the next one-half century the roster of punched card processing machines increased tremendously. The Remington Rand Corporation entered the field and, along with IBM, offered devices for sorting, punching, verifying, merging, collating, reproducing, printing, tabulating, and calculating. All of these machines, however, were dependent on electrical impulses to move mechanical components.

EARLY ELECTRONIC COMPUTERS

Although electromechanical machines performed yeoman service for more than 40 years, they were limited in both speed and flexibility. To move gears and manipulate cards takes time; and because each machine was designed to perform only a specific function, even more time was required to move the cards from one device to another to complete a given processing chore.

Christopher Evans writes that

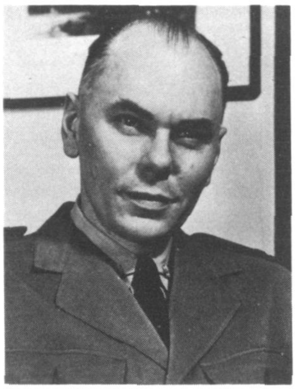
> in the 1930s a shift was beginning to occur, and the threads of the problem were being gathered together, quite independently, by a number of workers in various parts of the world. In the United States, large organizations such as IBM and Bell Telephone were at work. In England the thrust was coming from an individual, the mathematician Alan Turing, whose paper "On Computable Numbers," published in 1936, sent a jolt of enlightenment among the cognoscenti. In Germany, the threads were in the hands of a young engineer named Konrad Zuse, who had made up his mind not only to design a universal computer, but also to build one.

Konrad Zuse (1910-) IBM Archives

There is some debate as to who first actually came up with the idea of an electronic computer. As far back as 1915 James Bryce, a consultant to C.T.R., had at least suggested using vacuum tubes for data processing purposes; and George Stibitz of Bell Telephone Laboratories designed a computer using relay circuitry. But it appears that Zuse, an engineering student working on his doctoral thesis at the University of Berlin in Charlottenburg, was slightly ahead of the others in the building of a working machine.

In 1936 Zuse announced he was giving up his job as a design engineer to build a computer. Rather than fabricate components, he decided to use inexpensive, off-the-shelf parts. Although Zuse claims he was not aware of Babbage and his plans for a Difference Engine, there are some interesting parallels in design: both machines would have a memory, arithmetic section, output, and be capable of being programmed for any kind of mathematical job.

Zuse succeeded where Babbage failed. Over the next few years the German engineer built several successful calculators—each one an improvement over the previous machine. One of the calculators, the Z3, was used by a German aircraft manufacturer during World War II to solve wing flutter problems; another, the Z4, had some vacuum tubes to help it speed calculations for aircraft and missile design. Zuse was apparently well on his way to a full-scale computer when the wartime government decided to put its dwindling money and technical manpower into other areas of research. Some scientists today wonder what effect Zuse's computer would have had on the war's outcome if he had been able to complete it.



Howard Aiken (1900-1973)

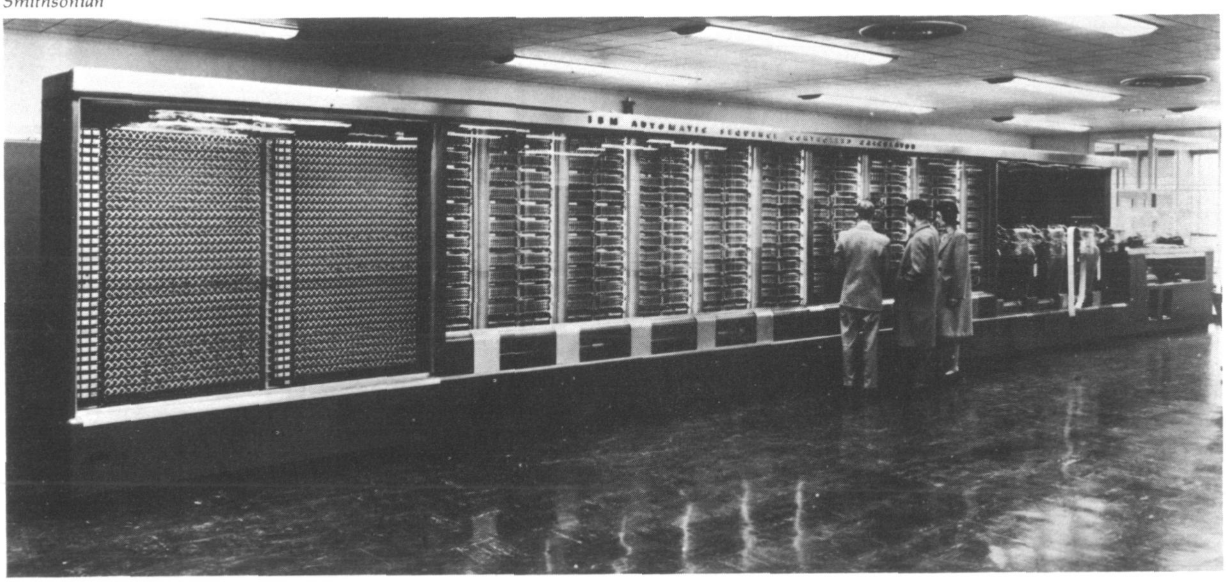
IBM Archives

About the same time in the United States, Howard Aiken of Harvard, with IBM's financial and technical help, built the electromechanical Mark I calculator which was essentially a linking of 78 individual adding machines and tabulators. The Harvard calculator occupied a gymnasium-size room and sounded, in the words of writer-physicist Jeremy Bernstein, "like a roomful of ladies knitting." The clicking sounds came from the rapid opening and closing of its 3,300 switches. During 15 years of use, the Mark I generated a huge amount of information that was used, among other things, for more accurate computation of the moon's orbit.

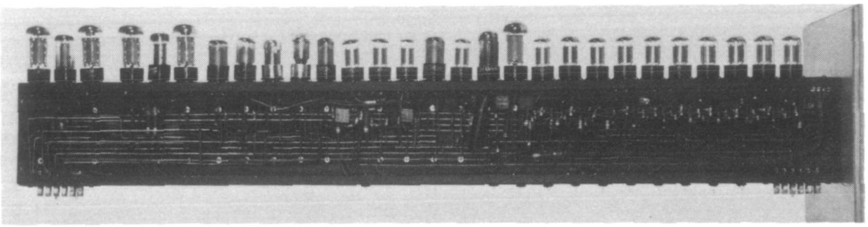
The first *true* electronic digital computer—that is, a machine using vacuum tubes for the generation and control of its electrical impulses—was the Electronic Numerical Integrator and Computer, more commonly called ENIAC. ENIAC was developed at the Moore School of Engineering of the University of Pennsylvania by J. Presper Eckert, Jr. and John D. Mauchly for the United States Army's Ordnance Department. It was a huge machine containing 18,800 vacuum tubes, and its inventors spent a good part of the first 2½ years just soldering the 500,000 connections needed for the tubes. ENIAC consumed huge amounts of electrical power, and its glowing tubes gen-

The Mark I calculator was built between 1939 and 1944 (with financial and technical help from IBM) by Howard Aiken of Harvard University.

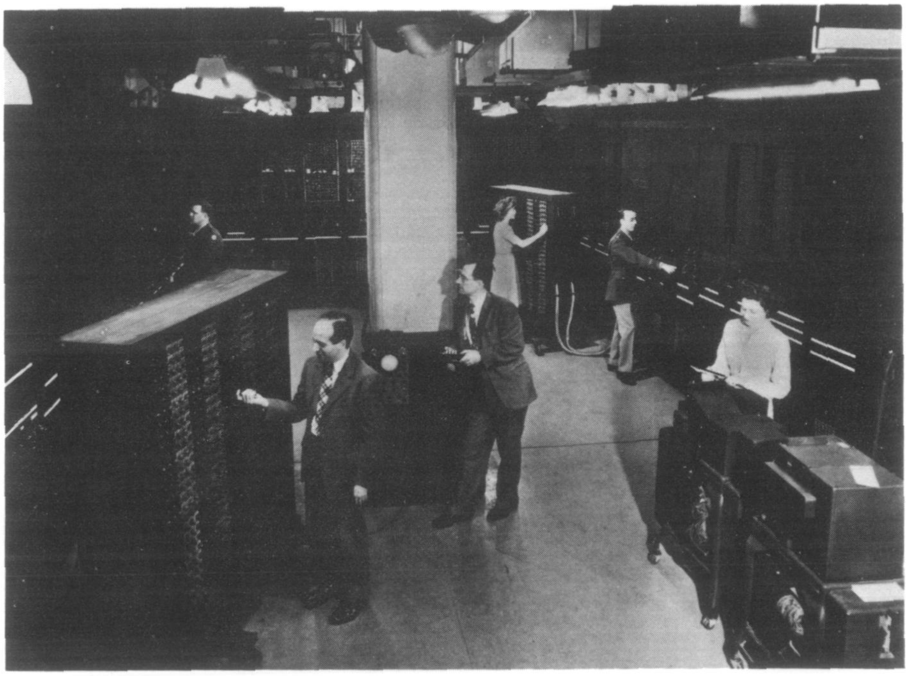
Smithsonian



erated a great deal of heat. It is said that every time the machine was turned on, three of its tubes burned out, and the lights in an area of Philadelphia dimmed momentarily.



Spare plug-in units such as this one were used to maintain the 18,800 vacuum tubes in the ENIAC electronic digital computer—a computer which required frequent maintenance.
Smithsonian



ENIAC, the first digital electronic computer, was invented by J. Presper Eckert, Jr. (left foreground) and John Maunchly (center foreground) at the Moore School of Engineering of the University of Pennsylvania in 1946.
IBM Archives

In operation, ENIAC could perform 5,000 additions per second. All internal functions were conducted by electrical impulses generated at the rate of 100,000 per second. And while this rate is barely a crawl compared with the speed of present-day computers, it was a tremendous advance over the capabilities of previous electromechanical machines.

ENIAC's primary job was to solve ballistics problems for the Aberdeen Proving Grounds in Maryland, where it saw service from 1947 to 1955. During this period Eckert and Mauchly developed still another computer called BINAC—a loosely formed acronym for Binary Computer—which eventually became the forerunner of Remington Rand Corporation's highly successful UNIVAC—Universal Computer—series.

Despite its far-reaching features, though, ENIAC lacked the one element that was needed to break open the development of computers. It did not have a stored program memory. Operating instructions for the machine were recorded by the manual placement of plug wires. As a result, once data were entered into the machine, they had to progress according to the paths laid down by these preset devices.

The Stored Program Concept

John Von Neumann (1903-1957)

IBM Archives

The critical advance into stored program computers resulted largely from the work of mathematicians involved in government projects to build machines intended primarily to solve military-related scientific problems. In the early 1940s John von Neumann was one of the permanent members of the Institute for Advanced Study (IAS) in Princeton, New Jersey. He also served as a consultant to the Army's Aberdeen Proving Grounds, as well as to the Atomic Energy Commission's (AEC) Los Alamos Scientific Laboratory.

It was his association with Aberdeen that led von Neumann to a chance meeting with Herman Goldstine, then a lieutenant in the United States Army serving as liaison for the ENIAC team. In the summer of 1944, the two men were waiting for a train on the Aberdeen railroad platform. Goldstine, a mathematician himself, recognized von Neumann and introduced himself. Von Neumann asked the Lieutenant what kind of work he was doing for the Army. Goldstine replied he was helping to build an electronic computer that could perform about 300 multiplications per second.

Von Neumann, as Goldstine later put it, was "galvanized." He immediately saw the possibilities of applying such a machine to computational problems in weapons design. "We got on the train together," Goldstine recalls, "and from Aberdeen to Philadelphia he pumped me for details."

Because of that chance encounter, von Neumann and his colleague at the IAS, Arthur W. Burks, entered into an active collaboration with Goldstine, Eckert, and Mauchly. Between 1946 and 1948, these men—along with several other scientists working at different institutions—published papers on computer design and program planning. And from this work emerged the concept of the stored program, a landmark idea that was translated into reality in a computer built at the IAS in 1952 under Von Neumann's direction.

Several other machines of note, designed or built during the 1940s, might qualify as the *first* stored program digital computer. Among them was the National Bureau of Standards SEAC built by Sam Alexander for the United States Army; the EDSAC constructed at Cambridge University in England; and the EDVAC at the University of Pennsylvania. All had features that allowed at least some program sequences to be stored in the same manner as data were stored.

Regardless of which machine came first, it was evident that computers could be extremely valuable tools for solution of sophisticated applied science problems. During the late 1940s and the early 1950s, the AEC funded at least five more pioneering systems which became part of computer lore: the MANIAC at Los Alamos; the ILLIAC at the University of Illinois; the AVIDAC at Argonne National Laboratory; the ORACLE at Oak Ridge; and the JOHNNIAC—named affectionately for von Neumann—at the RAND Corporation. During the early 1950s these machines contributed immeasurably to computer technology, and many of their individual innovations found their way into the general purpose computers of IBM and Remington Rand Corporation. "Perhaps our most important contribution," observes David H. Jacobsohn, who worked on the AVIDAC at Argonne National Laboratory, "was that we convinced industry there was an important scientific market for computers. At the time, however, we had no choice except to build these machines ourselves."

With the ability to store programs in the computer, the pace of development accelerated tremendously. Several United States companies were now building computers for the general marketplace. In Great Britain, which for a few years led the world in computing science, powerful data processing machines appeared in Cambridge, at Manchester University and at the National Physical Laboratory. A large English food corporation pioneered in the application of computer technology to commercial uses.

These early systems, however, had two distinct disadvantages: their enormous cost and the short life-span of their thousands of vacuum tubes. "You had to have a team of service engineers on hand at all times in order to keep up with burned out tubes," recalls an early user.

SOLID-STATE TECHNOLOGY

The vacuum tube, with its hair-thin wires sealed inside a fragile glass envelope, was a delicate and temperamental component at best. Even under optimum conditions, it had only a limited life-span. A large flow of electrical current was needed to get its filament hot enough to boil off electrons, but if this heat were not dissipated quickly and efficiently, the tube would burn out. Moreover, the whole unit was extremely vulnerable to damage or destruction from careless handling or external shocks.

Semiconductors

The solution to the fragility of early computers came from a long series of experiments with solid-state materials. As far back as 1874, a German physicist had reported a peculiar flow of electrical current in certain kinds of minerals. These minerals eventually became known as "semiconductors" because they conducted energy better than insulators, but not as well as conductors did.

Despite their unusual properties, semiconductors had remained, for the most part, just a laboratory curiosity. One exception came during the early days of radio when the semiconductor material, carborundum, was used as a "cat's whisker" detector—so-called because the two thin leads attached to the crystal resembled a cat's whiskers. The carborundum functioned like a subway turnstile, allowing electrons to flow more easily in one direction than the other. In this way, the crystal was able to convert the oscillating electron signal that came in through the antenna into a useable one-way flow of current.

When vacuum tubes were developed, however, they proved to be much more effective at this conversion. Moreover, tubes could also boost or amplify current. Because of these advantages, most of the practical work with solids was discontinued.

Nevertheless, many scientists were still interested in the unusual properties of semiconductors. The advent of quantum mechanics gave the first real understanding of how electrons in metals were free to move and conduct electricity. And in 1931 physicist H. A. Wilson published a theory of how electrons and holes in semiconductors and insulators gave those materials their electrical qualities.

These insights stimulated scientists at Bell Telephone Laboratories Incorporated to take a closer look at semiconductors as possible modulators of electrical communication signals. Early in 1940, for example, Russell S. Ohl, a staff member working with the semiconductor silicon, called several colleagues into his office to watch an unusual experiment. Ohl showed them a small piece of black silicon solid to which he had soldered two metal contacts. When light from a flashlight was shone on a narrow region near the middle, a photoelectromotive force of about 0.5 volts was developed. "I didn't believe what I saw," recalls a staff member, Walter H. Brattain, "until Ohl gave me a piece to work with in my own laboratory."

What Ohl and Brattain were working with was, in effect, the first *p-n* (positive-to-negative-flow) junction transistor. With the advent of World War II, however, there was little time to pursue the phenomenon any further. Although solid-state silicon detectors were employed during the war in radar devices (most of this work was done in England and at the Radiation Laboratory of the Massachusetts Institute of Technology), the detectors could rectify only certain high-frequency signals. Vacuum tubes were still required for amplification.

In January of 1946 scientific research on semiconductors was resumed in earnest at Bell Laboratories, with the directed goal of finding a solid-state amplifier. Among those who participated in this landmark work were William Shockley, Walter Brattain, and John Bardeen. Brattain wrote:

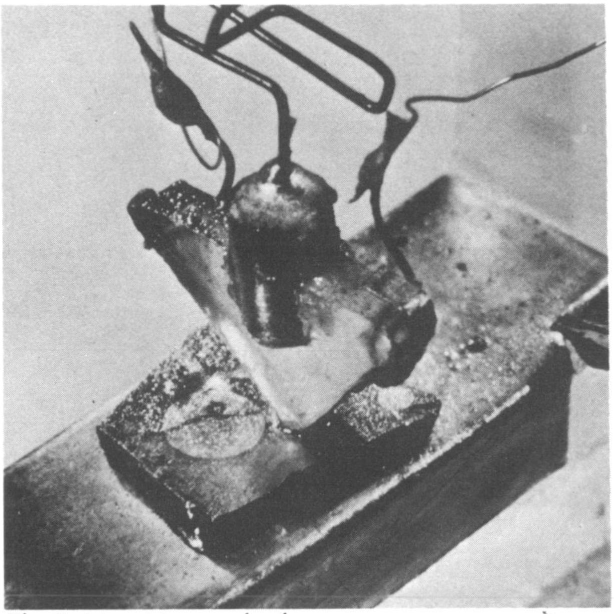
> At our first meeting, the group realized that in spite of all the work done before and during the war, we were still far from a real understanding. One reason was that copper oxide and other semiconductors on which early work had been done were very complicated solids. Silicon and germanium were the simplest, and the decision was to try to understand these first. Our work was directed toward a fundamental understanding of the problem though we were well aware of the technical importance of a semiconductor amplifier if one could be made.

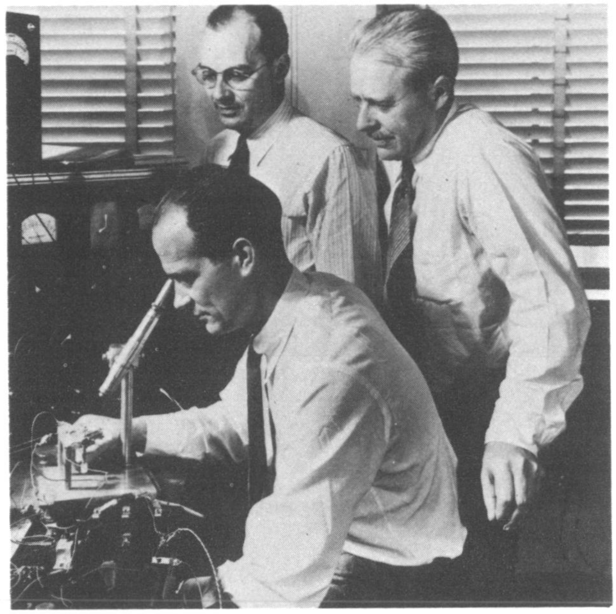
In December 1947, Shockley sent a casual note to a few colleagues inviting them to observe "some effects" the research team had come across during an experiment with a device that contained gold contacts and a germanium semiconductor base. "I hope you can break away and come," he added.

This may have been the understatement of the age. What Shockley's team demonstrated a few days later was the "transistor effect" by which they could control the movement of electrons in a semiconductor material through the influence of an outside electrical field. The action of the device, which was named a "transistor," was explained this way:

> The transistor's amplification process can be understood in terms of the discovery that the input point is surrounded by an "area of interaction." Within this

area the electronic structure of the semiconductor is modified by the input current. Now, if the output point is placed in this area, the output current can be controlled by the input current. This control of output current is the basic mechanism of amplification.





The first transistor was developed by William Shockley (seated), Walter Brattain (standing right), and John Bardeen (standing left) at Bell Laboratories. Although primitive by today's standards, this early apparatus demonstrated the transistor effect clearly enough to generate enormous interest in further research.

Bell Laboratories

The original transistor was called a point-contact transistor because it was essentially a wafer of germanium with two pointed-wire contacts located close together on one side. Shockley went on to work out the theory of *n-p*, *n-p-n*, and *p-n-p* transistors and also to design the junction transistor which, in many ways, was more effective and efficient than the earlier types. In a very real sense, the junction transistor sparked a technological revolution that has since changed the way people live.

But demonstrating an effect and actually producing quantities of workable transistors were two very different things. Early researchers found that it was almost impossible to predict how a given crystal would conduct current. Some allowed energy to move only in a positive direction; some permitted only negative flow; and some allowed two-way movement. Why the differences? Intensive research at several companies showed that crystals of germanium or silicon taken from nature almost always have a small number of foreign atoms locked among their molecules. Because these foreign atoms have either more or fewer electrons in their orbits, they change the electrical characteristics of the crystal. Thus, it was difficult to tell in advance how a given transistor would behave in an electronic device.

The people at Bell Laboratories—most particularly Gordon Teal and Ernest Buehler—began to study this problem and soon discovered they could "grow" single crystals of silicon in the laboratory. In this manner they could control the foreign atoms in the molecule and predict how the crystal would conduct current.

The key manufacturing process, however, turned out to be "zone refining," developed by Bell Laboratories' W. G. Pfann. As he explained it:

> The method of zone refining consists of slowly passing a series of molten zones through a relatively long ingot of impure solid. As a molten zone advances, impure solid melts at its leading interface, and purified solid freezes at its trailing interface. Each molten zone which passes through the ingot carries a fraction of the impurity toward the end of the ingot. The purification increases with the number of zone passes. Germanium purified by zone refining is probably the purest known manufactured material.

W. G. Pfann (left), inventor of the so-called "zone-refining" process, is shown here with his associate, J. H. Scaff, who is holding a large single crystal of germanium purified by this technique.

Bell Laboratories

In fact, crystals grown this way proved to be so pure that only 1 atom in 10 billion (equivalent to a pinch of salt in 38 carloads of sugar) was an impurity. The result was long, pencil-diameter "loaves" of semiconductor—usually germanium—which were then sliced into individual transitor pieces.

To control the electrical characteristics, Bell Laboratories and Texas Instruments, among others, developed several methods of "doping" the loaf with foreign atoms. One way to achieve the effect was to dope loaves with a few atoms of arsenic. Because arsenic has five electrons in its atomic configuration—one more electron than is found in germanium—the fifth particle was free to "roam" and produce *n*-type conduction. Doping with gallium, which has only three electrons, left "holes" or "escape routes" in the molecular structure which gave the crystal a *p*-type of electrical flow. Solid-state transistors could, therefore, be designed to perform most of the functions formerly handled by vacuum tubes.

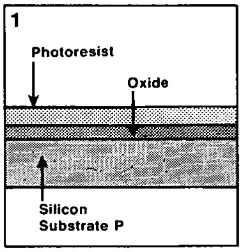
Integrated Circuits

Although transistors were originally designed for telephone switching applications, they arrived at an even more opportune time for the computer industry. The extremely high component failure rate in computers with thousands of vacuum tubes became a thing of the past. When computer manufacturers started replacing tubes with transistors, the failure rate fell from once every few hours to once every few days at most.

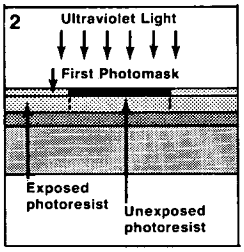
By the end of the 1950s, a second generation of computers had appeared, using germanium transistors as their basic active circuit element. At the same time, storage mediums for computers were improving rapidly—from tubes, to mercury delay lines and magnetic drums, to magnetic core technology (developed at MIT by Jay Forrester). In this latter technique—still widely used in computers—information is stored as magnetization in a tiny doughnut-shaped ferrite core. As a result, a large volume of instructions can be stored in the computer, making the system faster, more dependable, and more flexible.

All of these technological improvements served to make the computer a much more desirable tool. As *Time* Magazine reported, "No one took to the (new) computer more eagerly or saw its usefulness more quickly than the businessman." General Electric became the first commercial organization to acquire a data processing system; and many other major firms, more particularly insurance companies with their huge information handling needs, followed within a few years. Nevertheless, it was the military that played an especially significant role in early computer development. The huge IBM STRETCH system,

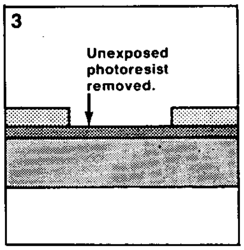
Doping a Silicon Chip



A wafer of silicon is oxidized and coated with photo-resist, an emulsion that hardens upon contact with light.



Ultraviolet light is beamed at the wafer and projects hundreds of tiny mask patterns onto the surface of the wafer.



The wafer is developed-washed in a special solvent, removing the photo-resist—except in areas struck by the ultraviolet light. Where the photo-resist has been removed, the oxide is now exposed.

for example, with 150,000 transistors capable of executing 100 billion instructions per day, was built in response to the Ballistic Missile Early Warning System's requirements for almost-instant data analysis and computation. UNIVAC III and several large transistorized systems from Control Data also were developed for equally sophisticated scientific applications.

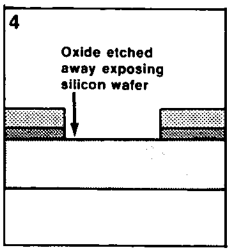
At first, manufacturers looked for clever ways of packaging transistorized circuits. One of the earliest featured transistors and printed circuit patterns wired together on cards which could be plugged into, or taken out of, a main frame. Thus, if a failure did occur, the faulty circuit could be replaced easily.

But even occasional failures were much too frequent for many users, particularly when the computer was a key element in a defense system. Analysis showed that most of the problems occurred in the interconnections rather than the transistors themselves: in the solder joints, at the plug-and-socket connections, at wire-wrap joints, and so forth. And the next developmental step was to find a way of minimizing or eliminating these troublesome spots.

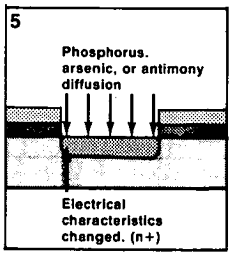
A new industry had developed in the United States in the early 1950s—the semiconductor industry. In 1956 William Shockley, co-inventor of the transistor, left Bell Laboratories to form the Shockley Transistor Division of Beckman Instruments in Palo Alto, California. A year later, eight of Shockley's best engineers left, and with the backing of Fairchild Camera and Instrument Corporation started a division which became known as Fairchild Semiconductor in the same area of Santa Clara County. Within a few years, so many new semiconductor corporations had been set up in and around Palo Alto that the region became known as "Silicon Valley." One of the redeeming features of semiconductor research, development, and manufacture was that a new firm could be started with relatively modest amounts of venture capital. Even as Fairchild Semiconductor prospered and grew, several of that company's best engineers left to start their own operations.

It did not take long for this activity to start paying off. In 1953, for instance, a Radio Corporation of America physicist applied for a patent on a circuit that could be fabricated on a single block of germanium; a year or two later several English scientists extended the concept; in 1958 an engineer at Texas Instruments succeeded in making an integrated circuit (IC); and a few months after that a team at Fairchild Semiconductor began work on what today are called microelectronic circuits.

Simply put, an integrated circuit is one in which all the active and



An etching bath removes the oxide in those exposed areas, revealing the original silicon wafer surface.



The wafer is then "doped" with the desired impurity by putting it into a diffusion furnace in the presence of an impurity vapor and baking it at a temperature of 900° Celsius for more than an hour. In this process of diffusion, the gaseous elements enter only those areas where the oxide was removed, altering the electrical characteristics of those areas.

passive elements are formed together on a small chip of semiconductor material. There are no separately installed joints, wires, or other parts that can break or become unraveled. Nevertheless, even an integrated circuit can fail—usually as a result of a manufacturing defect. The answer to this problem came from IBM engineers who designed a high-speed machine capable of testing the tiny chips before they are installed in a computer. Thus, once an integrated circuit is inside a data processing system, its failure rate is virtually nil.

Although the early ICs were formed on a germanium base, it soon became evident that silicon offered some distinct advantages. The reason lies in the basic process. Most integrated circuits today are made by the same kind of photoetching techniques used in metalworking industries. To start, engineers coat the surface of a semiconductor wafer with a thin photosensitive film called "photoresist." What this does, in effect, is to turn the wafer into a kind of photo contact paper. Then a mask or stencil of the circuit pattern that has been photographically reduced from a large drawing is placed over the surface and exposed to ultraviolet light. The light has the effect of either hardening or softening parts of the material. The softer material is washed away by a solvent, leaving the semiconductor surface open or exposed in the desired pattern. As a final step, doping impurities or metal for the interconnection lines are deposited on the exposed areas.

The photosensitive material, however, is insufficient by itself to prevent some dopants from depositing on other areas of the chip. Because of this, chemists started looking for a more impermeable material. Because its oxide is quartz, one of the hardest and most impenetrable materials known, silicon was an obvious choice. No other semiconductor, in fact, provided such qualities; silicon soon became the material-of-choice in chip manufacture.

Today the universal process for manufacturing integrated circuits (known as Silicon Planar Technology) is as follows: a thin wafer of silicon is oxidized and coated with photoresist. The resist, in turn, is developed to expose the circuit pattern; the wafer is then treated with hydrofluoric acid to dissolve the exposed areas of silicon dioxide, right down to the underlying base of pure silicon. The exposed silicon is finally doped with the desired impurity by baking it in a furnace in the presence of an impurity vapor—a process called diffusion.

Large-Scale Integration

Initially, single integrated circuits were replicated scores of times over on the starting wafer, and then the wafer was diced into individual chips. It soon became evident, though, that a much more functional device could be formed by interconnecting several of these circuits. The result was what industry people called Large-Scale Integration (LSI), in which a large number of circuits, each with a different function, are formed together on a single, ¼-inch-square silicon chip.

LSIs, which appeared in the late 1960s, gave the computer industry a huge boost. Because the first two generations of data processing machines had tremendous cost, heavy power needs, and huge heat dissipation problems, the only organizations that could even consider installing them were large government agencies, major scientific centers, and leading corporations.

With LSI technology, however, the size, cost, electrical drain, and heat generation of computers were reduced to the point where the systems came within reach of most medium-sized and some smaller companies. Not only did this mean a vastly expanded marketplace for computer manufacturers, but it also opened the gates for hundreds of entirely new firms. Among them were consulting organizations to help a company plan its data processing needs and install the equipment; "software" houses to write the special programs required by industry; peripheral equipment manufacturers to produce data storage facilities, terminals, and other "plug-compatible" devices which could be linked to a main frame; and a still growing number of smaller manufacturers making special function and miniaturized computers. Thus, what started as technology dominated by just a few firms soon involved hundreds of different companies.

Microprocessors

When Bell Laboratories announced the transistor back in 1948, no one could really foresee the revolution in electronics that would follow. Yet, by 1960, even the tiny transistor seemed bulky when compared with the integrated circuits and the large scale integration which crowded increasing numbers of components onto an ever-shrinking silicon square. Still another quantum jump in the technology of miniaturization can be witnessed today—the emergence of the so-called "miracle chip" on which both logic and memory circuitry are contained.

There is some question as to whether Texas Instuments or The Intel Corporation was first in developing microprocessors. The most widely repeated story, however, concerns a project which took place at Intel back in 1969. As *Time* Magazine described the events:

> Fresh out of Stanford University, where he had been a research associate, M. E. "Ted" Hoff was placed [by Intel] in charge of producing a set of miniature components for programmable desk-top calculators that a

> Japanese firm planned to market. After studying the circuitry proposed by the Japanese designers, the shy, self-effacing Hoff knew that he had a problem. As he recalls: "The calculators required a large number of chips, all of them quite expensive, and it looked, quite frankly, as if it would tax all our design capability."

To solve the problem, Hoff came up with a novel idea. Why not place most of the calculator's arithmetic and logic circuitry on one tiny chip of silicon? After wrestling with the design, Hoff and his associates at Intel were able to place nearly all the elements of a central processing unit (CPU)—the "main frame," as it is called in large scale systems—on a single chip. Finally unveiled in 1971, the one-chip CPU, now called a microprocessor, contained 2,250 transistors in an area barely 1/6 of an inch long and 1/8 of an inch wide. In computational power, the tiny microprocessor was almost as powerful as the original ENIAC, and performed as well as many IBM machines of the 1960s. It was, as the company was quick to advertise, "a new era of integrated electronics...a micro-programmable computer on a chip."

To most people, the word "larger" usually means "faster" or "more powerful." A large, eight-cylinder engine, for instance, produces more horsepower than a small, four-cylinder one. A mammoth steam shovel can probably dig out more dirt in one scoop than a person with a spade could move in many months. In the familiar mechanical world, a larger machine almost always does a job faster, more easily, and more economically than a smaller one.

In microelectronics, however, the reverse holds true. Chips are hundreds or thousands of times smaller than the vacuum tubes and transistors they have replaced. Yet everything about them is faster, more powerful, more reliable, and more economical. A desk-size computer today can often produce the same amount of work as a computer that once occupied an entire room. In computers, because a major limiting factor in the rate of computing is the time required to move electric signals within the machine, small is beautiful. As more components are packed onto a chip, signals travel shorter distances and calculating speeds go up; and as density and speed increase, computing costs go down. The result: many times faster processing at a fraction of the cost.

For those applications where small size is in itself a desirable end, the chip is made to order. Electronic wrist watches, versatile pocket calculators, portable foreign language translators, computer-controlled microwave ovens, automatically focused cameras, palm-size color television sets, automobile trip computers, engine monitoring computers—all of the new miniaturized computational devices on the market today—owe their existence to the thousands of transistors, memory cells, and passive elements crammed onto the surface of a silicon chip small enough to pass through the eye of a needle.

PROGRESS IN PROGRAMMING

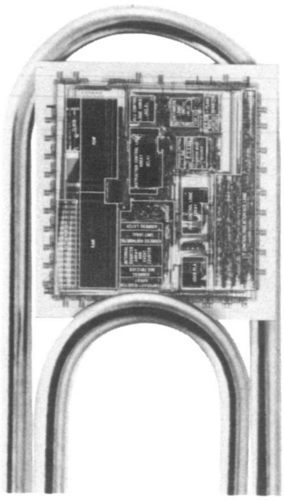
As circuitry got smaller and more reliable, an equally important change took place in the programming of computers. When data processing systems first appeared, most scientists needed an intermediary trained in programming to use them. Even with this kind of help, it took months or years to write a major program. The reason was that even the simplest problems involved the laborious and error-prone process of setting down long segments of instructions in the precise order needed by the machine. As a result, most organizations had to hire large—and expensive—programming staffs if they wanted to get their system "on the air" in a reasonable amount of time.

Grace Murray Hopper (1906-)
IBM Archives

The first major improvement came from the work of Grace Hopper, who started programming with Howard Aiken's MARK I computer group. She later served as a senior mathematician for the short-lived Eckert-Mauchly Computer Corporation and finally did some of her most productive work for the Remington Rand UNIVAC group. At Remington Rand, she developed the first practical "compiler" program which translated relatively compact instructions, such as "ADD C," into detailed binary code. More than anything else, her work showed that the computer itself could be used to do its own program writing, translating more or less standard English terms into binary numbers.

The next major advance came with the development of so-called problem-oriented languages. The first of these, still widely used, was FORTRAN, created by John Backus and his IBM colleagues in 1956. FORTRAN gave scientists, in particular, the ability to write their own programs in algebra-like expressions. Other math-oriented languages, such as ALGOL, PL/1, and APL, followed and had a profound influence on scientific computing. With ALGOL, for example, a complex program could be written in a few instructions instead of in several FORTRAN pages. Thus, programming became faster and "debugging" easier, with the net result of a drastic reduction in turn-around time between conceptualization of a problem and testing of the concept on a computer.

Today many of the instructions which formerly had to be written by a programmer are preprogrammed into a computer by its manufac-



The one-chip computer was developed for a variety of telecommunications applications. Here the size of the chip is compared with that of an ordinary paper clip.

Bell Laboratories

turer. Nevertheless, the art and science of getting a machine to do what is wanted still requires a skilled individual working with a particular programming language. Every computing system, from the largest distributed processing networks to compact stand-alone computers, consists of five elements: input, storage, processing, output, and control. To make sure a job is performed properly, the programmer must guide the system through all of these areas in a precise step-by-step manner.

While programming still requires a considerable amount of human skill and patience, an increasing amount of this work is now performed by the computer itself. Among the developments that have made this possible are the following: simplified programming languages and translators which take English-like statements and translate them directly into *0*'s and *1*'s; program libraries from which a user can select prewritten routines and piece them together to form all or part of a complex complete program; structured programming systems which allow an untrained individual to develop a program by answering machine-generated questions; preprogrammed memory chips which can "sense" certain internal activities and then supply the needed data management routines; and error-correcting codes, developed by Richard Hamming.

TRENDS IN COMPUTER USE

In the nineteenth century, Charles Babbage speculated that science would eventually "grind to a halt" because of a lack of calculating power. Instead, the opposite has happened; the pace of scientific development has picked up to the point where people are now being inundated with new information. This is true not only in science and technology but also in virtually every area of business and industry. Thus, one critical need today is to find ways of managing and using the ever-increasing information flow.

Clearly, computers are tools almost ideally designed for this kind of work. With their speed, storage capacity, display ability, and ease of programming, data processing systems are now finding their way into application areas not even dreamed of a decade ago. Here, for example, are just a few applications to illustrate some of the directions in which computers are now moving:

Scientific Information Storage and Retrieval

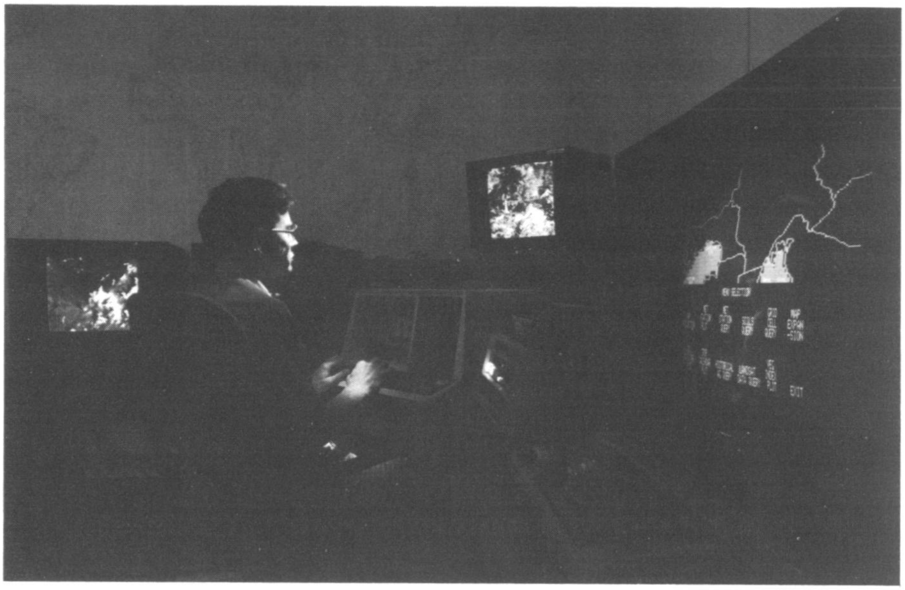
Computers are being used to store and to retrieve the huge amounts of scientific information on hand today. (It has been estimated that the amount of this information has roughly doubled during every decade of this century.) A major application in the scientific and technological fields has been the development of information storage and retrieval systems. Generally speaking, the information storage and retrieval systems being developed today feature huge data bases—with a given base usually confined to a specific scientific discipline—which can be reached through a variety of remote terminals. In most cases, the user needs very little training or computer knowledge to get at the material. "All it takes is a few key statements punched out on a typewriter keyboard," explains a cancer researcher at the University of California, Los Angeles, School of Medicine, "and the request goes directly to the MEDLARS (Medical Library and Retrieval System) computer in Washington, D.C." After that, it is only a matter of seconds before a title, abstract, or whole paper—depending on what is requested—is flashed back via the same channels, and either displayed or printed out by the same terminal.

Analysis of Satellite-Gathered Data

Computers are being used with a program designed to make better uses of the earth's resources. The program also is used to correct mapping information collected by satellite. As the National Aeronautics and Space Administration's LANDSAT satellite circles the globe each day, its cameras and other sensor devices map sections of the earth's surface and transmit digitized information to a land station. Ordinarily, such things as satellite roll, pitch and yaw, earth rotation, and sensor errors would make these digitized pictures very difficult to read. But the corrective "lens" of the computer's program is able to reconstitute each of its 115-mile-square pictures with remarkable clarity—even filling in sections that are missed by the cameras. Under certain conditions, even a single vehicle can be identified.

One key use for this satellite-collected computer-processed data is in agriculture. Through the program, the computer can help to identify acreage under cultivation by type of crop, can be used in the assessment of damage from floods or insect infestation, and, most significantly, can be used to predict crop yields. In 1979, for example, satellite data were used to estimate wheat production in the United States and Canada—a figure which turned out to be more than 95 percent correct when the actual crop was harvested. By the end of 1980 the same program was applied on a global scale to make predictions about world output of wheat, corn, cotton, rice, and soybeans.

Another important area for such systems, already being used to some extent by a few oil companies, is in cutting the time and effort



With the aid of LANDSAT imagery, a United States Department of Agriculture statistician is developing new mathematical procedures to predict crop yields.
USDA

involved in energy exploration. Using computer-produced pictures of seismic reflection patterns, geologists can spot formations in the earth which may contain oil or gas deposits. With this kind of information, they can then concentrate their land exploration efforts to increase chances of a find. (The importance of the computer in analyzing data to help locate oil and gas deposits is described in the chapter, "Seismic Exploration for Oil and Gas.")

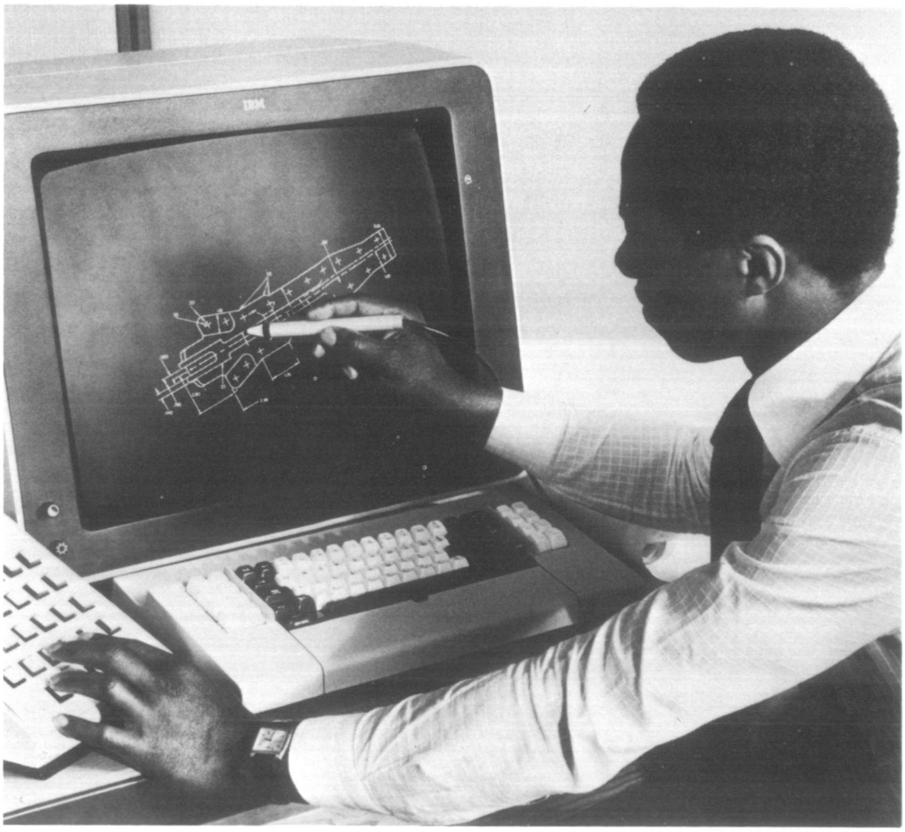
Other areas for satellite-computer applications include: land mapping, coastal zone management, monitoring water run-off from mountain snows, tracking icebergs, urban planning, selecting right-of-way corridors for pipelines, power lines, and highways, and gathering intelligence information. In the near future, satellites in stationary orbits over the oceans of the world will, through land-based computers, help to keep track of ship traffic.

Industrial Applications

In industry, goods-in-transit systems are already proving to be a rich vein for computers. Presently, many companies use computers to assist in the complex process of rating freight bills, improving utilization of trucks or railroad cars, and scheduling preventive maintenance for their fleets.

The greatest potential for savings, however, lies in scheduling goods directly from shipper to consignee rather than simply from yard to yard. Studies show that a loaded rail car moves toward its destination only about 11 percent of the time, leaving room for a tremendous amount of improvement. A mere 1 percent improvement, for example, could free up to 17,000 rail cars. Thus, computer applications which help to increase resource productivity are among the most important in industry today.

Another industrial application now clearly coming into its own is computer-aided design (CAD), sometimes referred to as engineering graphics. Since CAD came into existence about 10 years ago, most of the large automotive and aerospace companies have made it a fine art. What is significant today, however, is that many intermediate-sized manufacturers are now reaping product improvements and cost savings through CAD. One CAD application being used at a heavy equipment manufacturing firm, for instance, has eliminated up to 50 steps between the creation of a basic drawing and the actual cutting of a part on a machine tool. And every one of these saved steps can be translated into dollar terms. Computers can even be used to design themselves. As more complexity has been needed in very small integrated circuits, CAD has been used successfully to evaluate the optimal arrangement of subcircuits and their connections and interconnections.



Computer-aided design (CAD) has been of great service to the engineering-design industry by reducing the costs of production through the elimination of drafting steps.
IBM Archives

Commercial Applications

Although business has long used computers for accounting and recordkeeping functions, the accelerating trend today is toward electronic, "paperless" offices. Many firms now use computers for text entry, information retrieval, composition, and for electronic transmission of letter-images, often via satellite, from computer to computer within a company. In addition, many large legal organizations now use computerized systems for both storage and retrieval of pertinent legal data.

Point-of-sale computer applications, such as the terminals used in retail and grocery firms for credit card transactions, check cashing authorization, debit card activities, and automatic laser-beam scanning of bar codes, are gradually replacing the traditional cash register. Such systems provide productivity improvements, timely management information, and better control over store operations in one recording step.

Yet the trend is not limited to these areas alone. Drug stores are now installing pharmacy systems to improve the process of filling prescriptions. When a customer purchases a prescription drug, the computer provides information such as side-effects, expiration dates, prices, and a receipted record for the customer's tax purposes.

Medical Applications

Hospitals and physicians are finding more and more uses for computers. Computerized Tomography (CT) scanners now use computers to intensify and analyze cross-sectional and longitudinal X-ray pictures of the body. (CT scanners are discussed in more detail in the chapter, "X Rays for Medical Diagnosis.") Computers also monitor and signal events in the Intensive Care Unit (ICU), help to provide rapid and accurate readout of laboratory tests, and keep track of patients' progress from the time they enter the hospital until long after they have been discharged.

Scientific Laboratory Applications

Today there is hardly a scientific laboratory that does not make use of computers. Minicomputers, and even microcomputers, are often integral parts of such sophisticated experimental machines as Nuclear Magnetic Resonance (NMR) and Electric Paramagnetic Resonance (EPR) apparatuses. In the most advanced instrumentations, data are handled from their source to their ultimate disposition entirely by a computer. The human researcher intervenes to direct the course of an experiment, interpret results, and generally manage the laboratory system. But all of these functions are enhanced to some degree by a computer.

Finally, the computer is increasingly being used in the home. Originally, applications of computers away from the work place were limited primarily to a handful of engineers, scientists, students, or sophisticated people in business who were obliged to do a lot of work at home. With the advent of the microprocessor, however, all sorts of convenience applications—from home recipe planning to energy savings systems—are becoming apparent.

QUALITY OF LIFE

Despite the vital nature of these information management functions and the usefulness of the devices made possible by microprocessors, the very mention of computers and chips often elicits only a negative response. Some people blame computers for many of the more unpleasant aspects of modern life: billing errors, junk mail, commercial crime, job dislocation, and threats to privacy and individuality. These problems should be viewed from the perspective of a technology that is still growing, if not exploding. Although the victims of computer technology gone awry are often told, "The computer did it," it is actually the people who program and run computers who create most of the problems for which their machines are blamed. "Computers do only what they are told," says a data processing scientist, "and not an iota more."

In the final analysis, it should be recognized that technology creates new products and services and, therefore, new job markets. Already, microprocessors have opened a home computer market; they have been responsible for scores of new calculator-type games and many different electronic children's toys; they have made possible many new miniaturized communications devices; and they are the keystone of a still-growing digital watch industry. "Nevertheless," says Bell Laboratories microprocessor chief Lee Thomas, "the most exciting applications won't come until the kids who are still in high school and have grown up with pocket calculators and home computers become the engineers of the 1980s and 1990s." Thomas also feels that if American business fails to exploit this capability, a potentially enormous market could slip through its fingers.

THE FUTURE

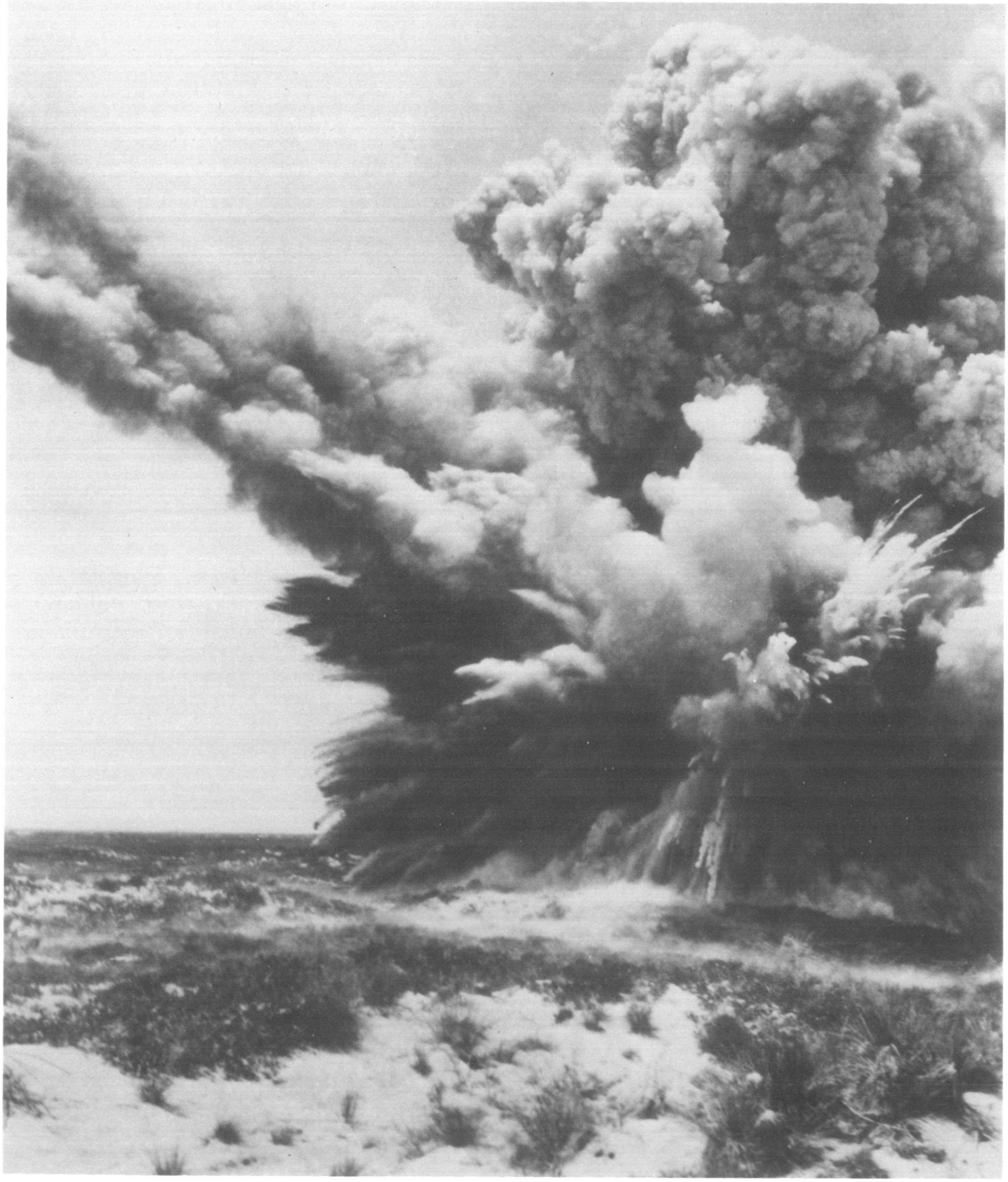
Despite the dramatic developments that have taken place in recent years, the world is still in an early stage of microelectronic technology. And from the perspective of more than 30 years of progress, a few conclusions stand out.

First, most of the devices and components used in computers today did not result from accidental discovery, but from intensive, mission-oriented research programs. In this field, typically, scientists have deliberately set out to find alternatives to components already in use.

Secondly, the unparalleled advances in semiconductor and computer technology during this period were the products of interdisciplinary teamwork. The complexities of integrated circuits—not to mention the complexities of putting a computer together—required the correlated efforts of specialists in materials, components, circuits, and systems. And the job of moving the circuit from a drawing board to its place in a computer required still another team of highly trained people.

Finally, applying computers to an ever-expanding range of jobs is dependent on people whose skills were not even known 30 years ago. These people—among them programmers, systems analysts, and even a new breed of sales representatives—have all contributed to the success of this new, yet still growing industry.

Few technologies in history have come so far so fast. Thirty years ago people were looking forward to the "nuclear age." However, the computer's spectacular growth—in numbers, in power, and capability, in the variety of things it can do, and in the economy it provides when doing these things—has brought about the "computer age." What is more, it is a safe bet that this new era will be around for a long time to come.



Seismic Exploration for Oil and Gas

In a Connecticut Veterans Administration hospital, a man who has lost his eyesight listens carefully to stereo recordings of someone walking through hallways, in and out of rooms, and along city streets. His instructor points out important clues in the quality and loudness of the sounds. With this help, the sightless patient will be able to orient himself; he will soon be able to distinguish differences in the reflected sound of his footsteps or of his cane taps. He will be able to tell if he is walking along a corridor by the speed with which the echo of his footsteps returns from the nearby walls; he will learn to distinguish differences by the change in sound quality when he walks into an intersection of corridors; he will learn to listen to those echoes for clues to the size, the shape, and even the furnishings of a room. In other words, the blind patient is being trained to sense and process information about his surroundings based on his experience of the reflection time of sound waves through air.

Someone standing on the surface of the ground who needs to learn about the structure of the earth 100 feet, 1,000 feet, or even 1,000 miles below is like the blind man. Although a sightless person could conceivably walk around the room touching walls and feeling surfaces, a similar technique is clearly impractical for a geologist or geophysicist trying to learn about the structure of scores of cubic miles of earth. One of the most useful methods of the modern geophysicist is analogous to the technique that the patient at the Veterans Administration hospital was learning. The geophysicist, too, can learn about the world by interpreting the behavior of waves—in this case, seismic waves as they travel through the earth.

setting off explosions such the one shown on the oppo- page, geophysicists can gain ormation about the composi- n of the various strata below surface of the earth. Such ormation helps them locate as where oil and gas may have umulated.

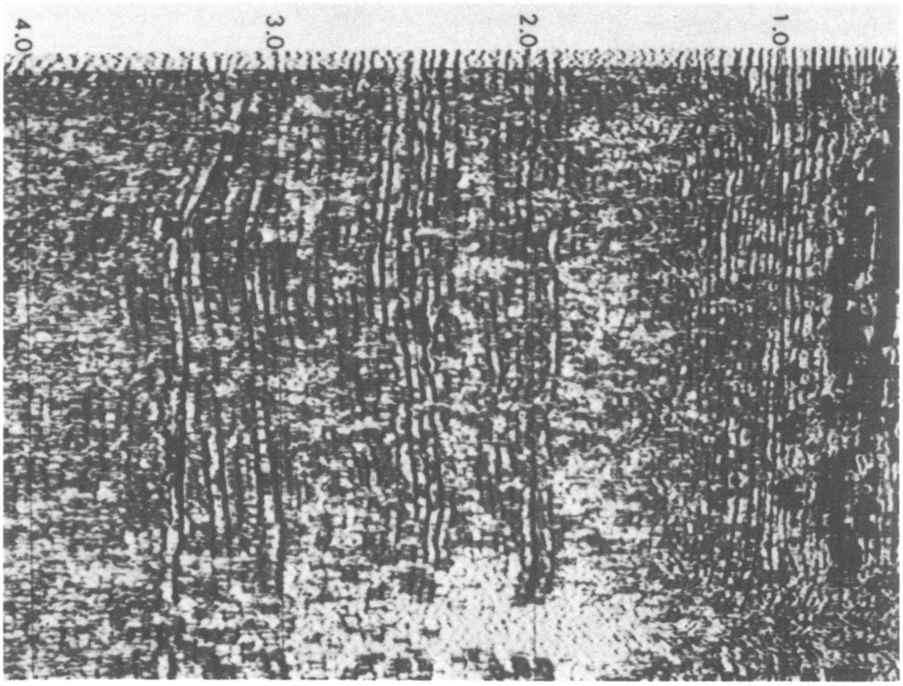
Scott Petty & Geosource, Inc.

The technology that allows the geophysicist of today to see below the surface of the earth is based on contributions of scientists from a broad variety of scientific fields. Many of these scientists, such as Leonardo da Vinci, Isaac Newton, and Christian Huygens, were trying to satisfy their curiosity about how things work, how the earth's surface came to be, or how natural phenomena, especially light and sound, behave. Da Vinci's geological observations, Newton's conception of the physical universe, and Huygens' unravelling of some of the mysteries of the behavior of light, were combined hundreds of years later by geophysicists trying to develop versatile and reliable methods of locating oil and gas deposits. In the work of many of the giants of science, others saw techniques and concepts that could be shaped and adapted to search for oil and gas and even to explore the inner structure of the earth to its core.

BASIC PRINCIPLES OF SEISMIC EXPLORATION

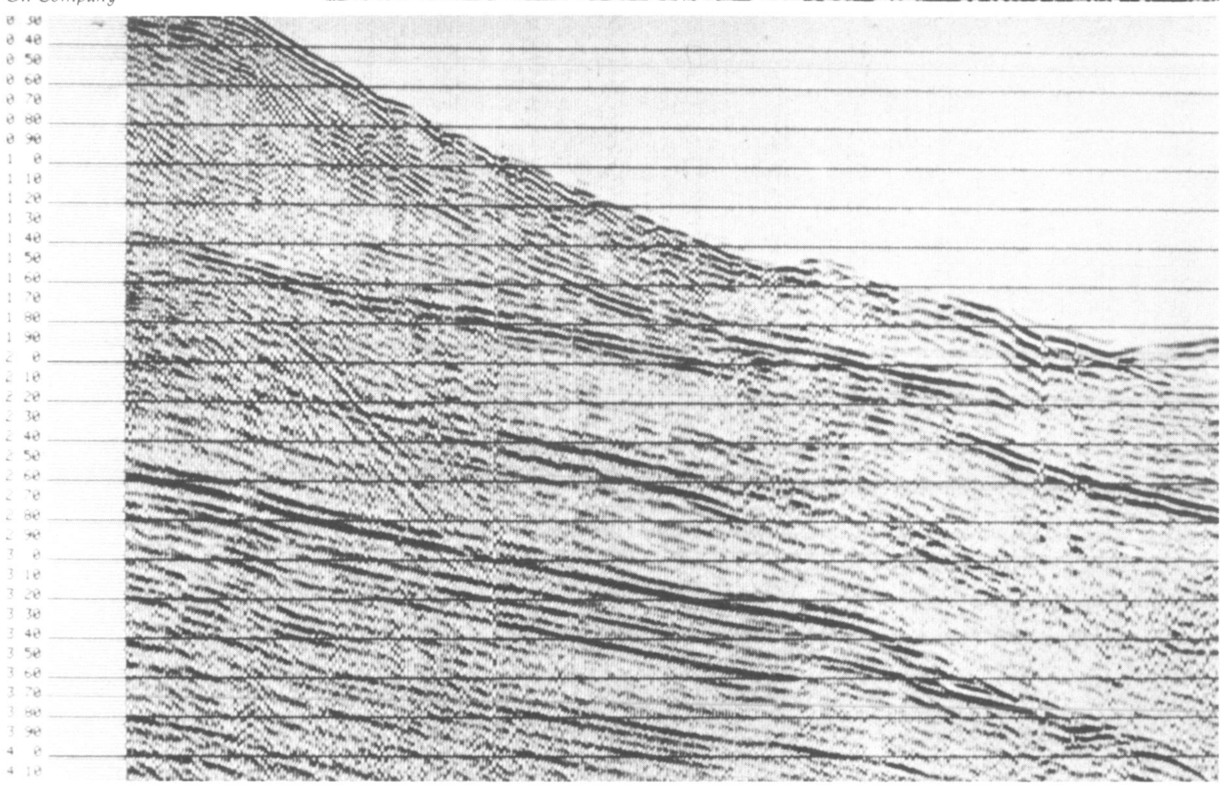
Anyone who has felt the ground tremble underfoot has sensed clearly that vibrations travel through the earth. Whether it is a heavy truck, a subway train, an explosion, the demolition of a building, or an earthquake that causes the ground to tremble, they all have one thing in common—they send waves (vibrations) traveling through the earth. More than 400 years ago, English soldiers besieged at Exeter recognized and took advantage of this phenomenon. Concerned that their enemies might breach their defenses by tunneling under the walls, the soldiers set out pans of water and watched them carefully for surface ripples caused by the impact of shovels digging below. The modern seismic method of exploration for gas and oil takes advantage of the same principles of detecting vibrations, but modern technological progress has produced vibration-detecting equipment far more sophisticated than pans of water set out on the ground.

Today, application of the seismic method to look for oil or gas uses explosives or other energy sources, scores of specially designed geophones, magnetic tape recorders capable of recording up to 1,024 channels of sound simultaneously, and computers as powerful as those used in the space program. While it was important for those soldiers at Exeter to know if their enemies were tunneling, they needed to know only whether or not there was vibration. Interpretation was simple—if there were vibrations that could not be explained otherwise, the enemy was at work below. Today's geophysicist is looking much farther and deeper than those sixteenth-century soldiers. To reach useful conclusions, modern geophysicists use far more precise data. Still, the



Recordings made from modern seismographs enhance the ability of geophysicists to make interpretations.

Mobil Oil Corporation/Marathon Oil Company



interpretation of the stacks of computer printouts and graphs is simply that—an interpretation. Seismic exploration does not indicate where oil is, only where it is likely to be. And in the case of seismic prospecting for oil, a decision to drill can mean millions of wasted dollars or it can mean the discovery of significant deposits of oil, the development of natural resources, and corporate profits—an awesome gamble to base on what is no more than an "educated guess." For the guess to be as educated as possible, the geophysicist needs every scrap of information and every clue available.

Formation and Accumulation of Petroleum

The ways that petroleum formed and accumulated have dictated the kinds of clues geophysicists seek with their sophisticated equipment. Most geologists today believe that petroleum is formed from the remains of plants and animals which lived millions of years ago in or along the coasts of the shallow seas, swamps, and lakes that covered many parts of the earth. When these organisms died, their remains eventually sank to the bottom and were covered by sediments, such as silt, shale, or limestone. As the sediments and organic matter were buried deeper and deeper, they were subjected to increased temperature and pressure. The sediments gradually formed rock, while bacteria, heat, pressure, and diverse natural processes, which are still not totally understood, transformed the organic material into oil and gas. The sedimentary rock in which the oil and gas were formed, commonly mudstone or shale, is called "source rock." As layers of sediments continued to accumulate under the water, the newly formed oil and gas were subjected to increasing pressure. The pressure caused the oil and gas to move into nearby porous sedimentary layers called "reservoir rocks" which were under less pressure. Within the same formation, oil and gas may also move upward because they are less dense than the water which fills the spaces in the rocks.

The reservoir rocks, no matter what their composition—sand, sandstone, or limestone—have two very important properties: they are porous and permeable. The oil moves into the pore spaces within these rocks in much the same way that water poured into a glass filled with sand finds its way into the empty spaces between the grains. The more porous the reservoir rock, the more oil and natural gas it can hold.

In addition to the source rock where the oil and gas are formed and the reservoir rock in which they can be stored, a third condition is necessary for significant oil or gas accumulation—the presence of a trap. What happens when oil or gas is not trapped can be demonstrated with a simple oil lamp. When a wick is put in the oil reservoir of the lamp, the oil saturates the wick. The longer the wick, the longer it takes for the oil to reach the end. Eventually, though, the wick will draw all of the oil from the lamp, resulting in a damp wick and a dry reservoir.

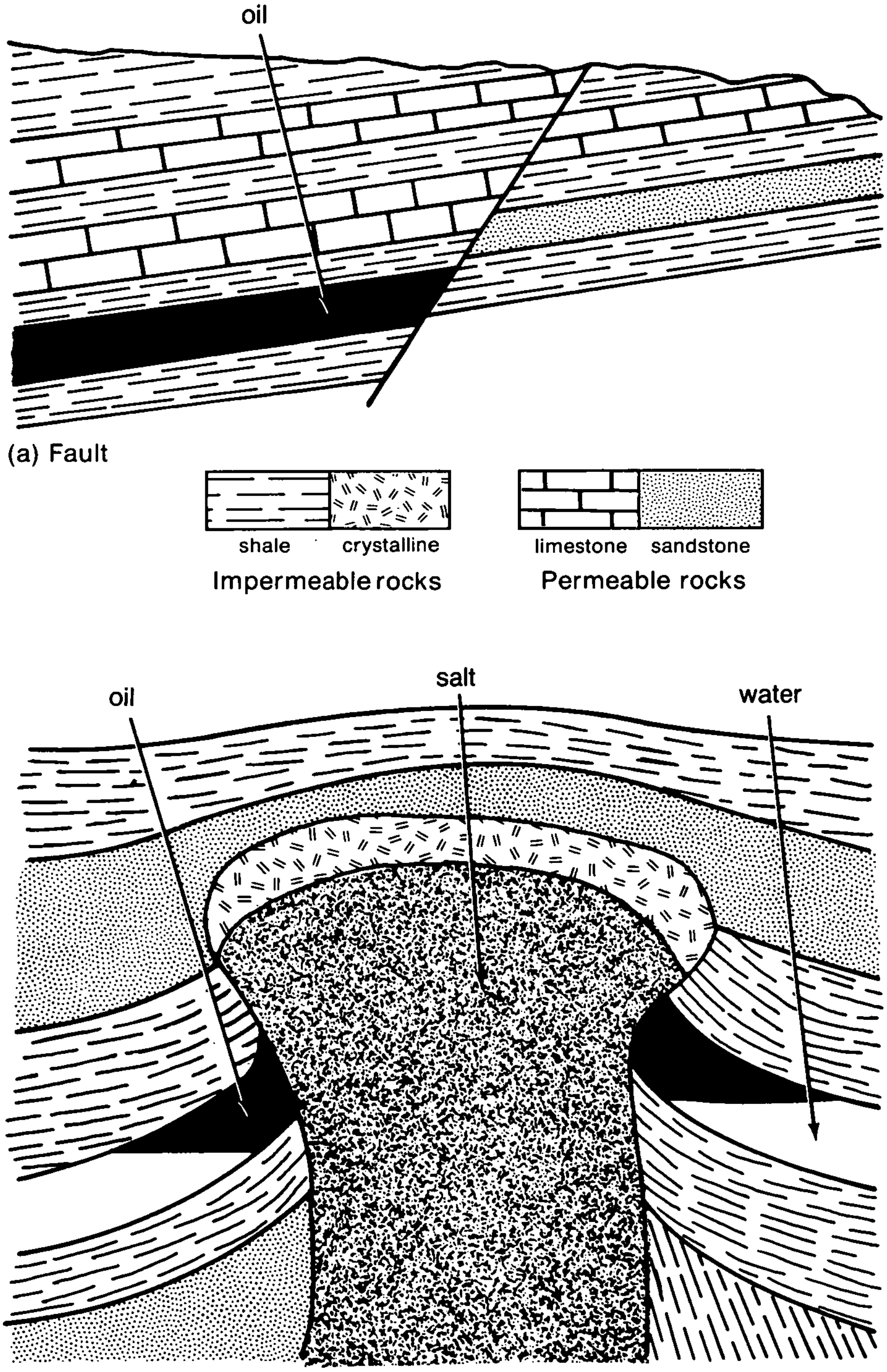
With crude oil in the earth, the porous rock is both the reservoir and the wick. The oil, unless blocked, migrates upward and laterally through the porous rock layers until it is dispersed so thinly that, for all intents and purposes, it disappears. It may even reach the surface where the lighter fraction may escape into the atmosphere, as in the well-known case of the La Brea tar pits. Luckily, oil can be trapped in the reservoir rocks if the adjacent rock is nonporous. But how can this happen if the sedimentary layers are horizontal?

Almost 500 years ago Leonardo da Vinci glimpsed part of the answer. The fossil seashells that he found in the Alps convinced him that the surface of the earth had undergone gigantic changes. During the long periods of time that it took for the formation of oil, segments of the earth itself shifted. Huge areas of the planet's surface rose and fell. New bodies of water appeared, and the bottoms of ancient bodies of water became high plains, deserts, and sometimes even mountains. During this time, the earth's crust buckled, cracked, and folded, and faults developed. This geologic activity millions of years ago set the stage for modern technological society by forming traps for the accumulation of oil and natural gas.

One common example of such a trap is the salt dome. The oil and gas around a salt dome were formed in an organically rich sedimentary layer. Buckling in the earth's crust, probably caused by salt pushing upward, deformed the original horizontal layering of the sedimentary material. This action of the salt dome created fractures in the distorted layers, trapping oil and gas around the dome. The gas, less dense than oil, flowed to the top of the trap. Today, the geophysicist who locates a salt dome knows that this is the kind of structure that *might* be associated with oil. Only a successful well will tell for sure.

Breaks in the earth's crust, or faults, are another common form of hydrocarbon-trapping structure. Although the break can be in any direction, the one commonly associated with oil entrapments forms when the earth shifts vertically and causes a permeable layer to lie next to an impermeable layer. Again, the story begins with the deposition of organically rich sedimentary layers. The organic material begins to be transformed into oil and natural gas. As the layers shift in their relation to each other, a fracture with displacement—a fault—may be formed.

PRINCIPAL TYPES OF OIL TRAPS



(b) Salt dome

Typical petroleum traps. (a) Fault. (b) Salt dome.

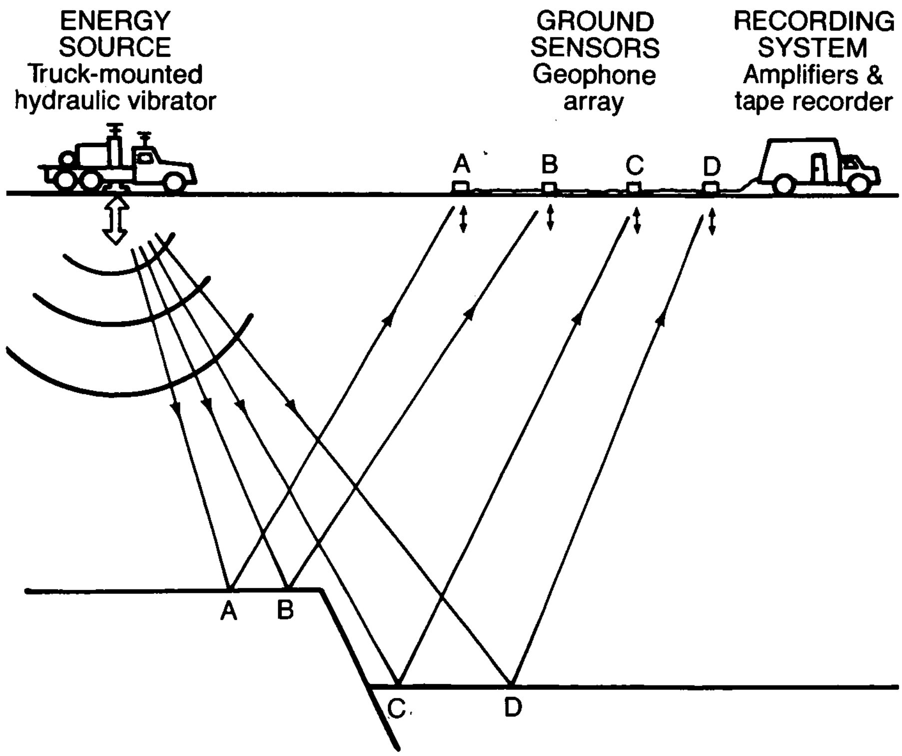
As petroleum is formed, it tends to move generally upward along permeable beds. In a particular circumstance, for example, perhaps the oil flow can move only toward the right and up until it reaches the fault itself. Because the right side of the fault consists of an impermeable layer, the oil and gas must stop. A new reservoir is formed.

There is an almost infinite variety of traps that might contain oil. What is important to remember is that while accumulations of oil are found only in traps, not all traps contain oil. The fact that a geologic structure could trap oil does not mean that it did. The proof is in the drilling.

Behavior of Seismic Waves

Before any kind of interpretation of information gained from seismic waves is possible, however, geophysicists need to understand the behavior of these waves as they travel through the earth. Today it is known that as a seismic wave travels through the earth, a part of it is

Seismic reflection profiling using the Vibroseis® technique



bounced back each time it encounters an interface between different layers of rock. This is a little like what happens when ripples set off by a pebble in a pond strike an object in the water. Someone looking at the water closely will see smaller ripples that have bounced off that object moving away from it. In the case of a seismic exploration, the waves reflected from rock layers bounce upward and travel back to the surface where sensitive detectors, called geophones, record them. The geophone is similar to a microphone. It "hears" the vibrations in the ground and transmits them to a tape recorder.

In the same way that the captain of a ship learns the depth of the water by learning how long it takes a "ping" or sound wave to reach the ocean bottom and return, geophysicists measure the time it takes the seismic wave from a small explosion to bounce back to their detectors. But, in dealing with a single and fairly uniform element—water—the ship's captain has a comparatively easy task. Sound travels through water with a known speed. Distance, then, is easy to calculate. Once the speed of sound in water is known and the time it takes for the echo to return is measured, distance can be calculated. Within the earth, however, the application of the same principles is far more complicated for a number of reasons.

In 1845 an Irish engineer, Robert Mallet, tried to discover the speed that sound travels through the earth. He got the wrong answer, but his experiment was so ingenious that his methodology made the largest single contribution to the seismic method in the nineteenth century. He buried a charge of gunpowder connected to an electric detonator and placed bowls of mercury various distances from the explosion. A spotlight on the mercury illuminated the liquid so that with a telescope he could see ripples on the surface when the seismic wave reached the bowl and made it quiver. To time the wave he used a clock which was started electrically and simultaneously with the explosion. He could measure the distance on the ground between the site of the explosion (today called the shotpoint) and the mercury (corresponding to today's electronic sensors). By measuring the time it took for the seismic wave to reach the bowl, he could determine the speed of the wave. Mallet also made the very important observation that different velocities were derived from different surface geologic materials. The observation that the velocity of the seismic wave was dependent on the sort of rock through which the wave was traveling was to be of immense importance to twentieth-century geophysicists. For example, it was discovered that seismic waves traveled through rock much faster than through soil. Because of the comparative insensitivity of the mercury, the actual velocities Mallet calculated turned out to be too low. Nevertheless, his work was a success. In the middle of the nineteenth century Robert Mallet had devised all of the basic elements of the seismic method. He had also discovered that the speed of a seismic wave was related to the material through which it was traveling. Modern developments with their more sensitive and complex electronic components are, in a real sense, refinements of Mallet's pioneering work.

PROBLEMS ENCOUNTERED WHEN USING THE SEISMIC METHOD

From the beginning, the problems of "seeing" underneath the surface of the earth confused the data. Mallet did not know that his mercury detectors were giving him inexact information because the mercury was not sensitive enough to signal the arrival of the first weak vibrations arriving at the bowls. Today, with all the advances in physics, geology, and electronics, geophysicists still have to face an impressive array of problems that grow out of the behavior of seismic waves.

Velocity

The first problem facing the geophysicist is that, unlike water, the earth is made up of a variety of materials. In water, sound travels at approximately 5,000 feet per second, varying only slightly if the water is salt or fresh. But in the earth, sound travels anywhere from 1,500 feet per second in some unconsolidated muds, clays, and sandy materials, to as fast as 25,000 feet per second in some dense, deeply buried rocks. Geophysicists refer to the speed of sound through a material as the "seismic velocity" of that substance. Seismic velocity is determined by the solid state crystal structure, the density, the amount of pore space, and the fluid or fluids filling the pore spaces of the material.

In a sense, the seismic velocity of a rock layer is a signature identifying the type of rock. That, in turn, can be an indicator as to the likelihood of the existence of the proper conditions for the presence of oil beneath the surface.

Refraction

A second problem is also related to speed. Under most conditions, when a wave, including a seismic wave, changes speed, it also changes direction; it bends. Physicists refer to this as a change in the direction of propagation or "refraction." Light waves behave the same way;

when a ruler is put into water at an angle to the surface of the water, the ruler appears to be bent at the point where it enters. As the light waves travel from one medium, water, to another medium, air, they change speed; they are refracted if the path of the light wave is at an angle to the interface. The ruler appears bent. Since the layers of the earth are made up of rocks having different seismic velocities, seismic waves traveling through them are often bent too. It is important for the geophysicist to understand how much bending takes place. That information will give important clues to the velocity of the rock layers. These clues are extremely useful in decoding what sorts of rocks and what formations make up the area under investigation.

Analysis of refracted waves was the first seismic method employed to explore for salt domes and the oil which might be associated with them. The technique takes advantage of the fact that the seismic velocity of salt usually is considerably higher than that of the sedimentary rock beds surrounding the dome. Stated simply, if the seismic wave travels from the explosion to the detector a lot faster than it should if it were traveling through sand or shale, a salt dome and, better still, oil may have been found.

Reflection

Although the word sounds somewhat similar to refraction, reflection refers to a very different wave quality. If refraction is thought of as bending, reflection is like bouncing. To understand reflection, it is useful to think of what happens when a ball is thrown against a wall. A ball thrown directly at a wall bounces directly back. If the ball is thrown at an angle, the angle of the bounce is directly related to the angle at which the ball hits the wall.

Seismic reflections behave in a similar fashion. The seismic bounce takes place when a wave moving through one kind of rock hits a layer of another type of rock which has a different seismic velocity. It is the contrast between the seismic velocities in different types of rock that is the "wall." The difference in seismic velocities of two adjacent rock layers, for example a shale and a sandstone, is usually not very great. So the reflection property of sound waves means both good news and bad news for the geophysicist.

The bad news first. The amount of energy reflected back toward the surface is not very great. Typically, less than 1 percent of the downgoing energy gets reflected back to the surface at a single interface because so little of the seismic wave is reflected. The geophysicist needs very sensitive instruments to record these weak signals.

The good news is that this same lack of reflection means that most of the energy continues to travel farther down into the earth for reflection from deeper interfaces. It means that the seismic method can usually be used to map the subsurface through the thousands of feet of rock built up over eons of the earth's history all the way to the base of the earth's crust. This method can also be used to obtain information about the structure of the earth as deep as its core, because the energy that continues downward may also be "refracted" as it travels from one deep layer to another. At deeper levels, however, the data become less precise because of the dissipation of energy.

Noise

Seismic waves from an explosion travel in all directions from their source. The geophysicist wants to detect the refracted and reflected waves returning to the geophones on the surface. Unfortunately, the source of those waves, the original explosion, creates a lot of additional unwanted seismic waves that sometimes overpower the much weaker signals coming back from beneath the surface. The geophysicist refers to these unwanted signals as "noise."

The "snow" on a television set is electronic video "noise." This is something like the unwanted sound waves the geophysicist detects. When a channel with a better picture, or signal, is chosen the noise is reduced. A better antenna, say one on the roof rather than "rabbit ears" on the set, enhances the ratio of signal-to-noise, and the television picture improves. Fine-tuning might filter out even more of the "snow," improve the signal-to-noise ratio, and result in a better picture.

One common way for the geophysicist to improve the signal-to-noise ratio is through proper placing of the explosive. Often, the top layer of soil is made up of a variety of materials including rocks and boulders mixed with sand, decayed organic materials, and other substances. This layer, called the "weathered zone," produces confusing seismic information. By drilling through it to the layer below and placing the shot there, the seismic explorer can minimize the effect of this layer on the signals coming back to the geophones and reduce the noise—in effect, getting rid of the "snow" to get a better picture. Many of the significant improvements in the accuracy and utility of seismic exploration for oil and gas have been in this area.

Resolution

A fifth difficulty for the geophysicist using sound waves to explore the earth's substructure is that the distance from one rock layer interface to the next is usually quite small, often less than 50 feet. Reflected from adjacent interfaces, waves traveling at thousands of feet per second—thousands of miles per hour—overlap each other as they travel back to the geophones on the surface. The time differences in the arrivals of the waves are very small. This can make it quite difficult to map any single interface. To go back to the example of the television set, reflections from nearby buildings can sometimes be so strong that the "ghosts" on the tube overlap each other so that the lettering on the screen cannot be read or the picture may not be recognizable at all. In such a case, the signal from the television station's transmitter is going not only directly to the antenna but also to nearby buildings where it bounces perhaps to the antenna or perhaps even to another building, all before reaching the set. Because of the differences in distance, a part of the television station's signal takes a little more time to get to the set. The television set has trouble making sense of such a signal and decodes it as a "ghost" rather than as the picture originally transmitted. The television viewer usually considers this ghost a minor problem, jiggles the antenna, adjusts the controls, and decides to "live with it." But geophysicists, charged with the responsibility of making decisions involving hundreds of thousands of dollars, cannot afford to have such an attitude. They need all of the resolution possible. Resolution is one of the major problems of seismic research that modern technology has tackled with considerable and continuing success.

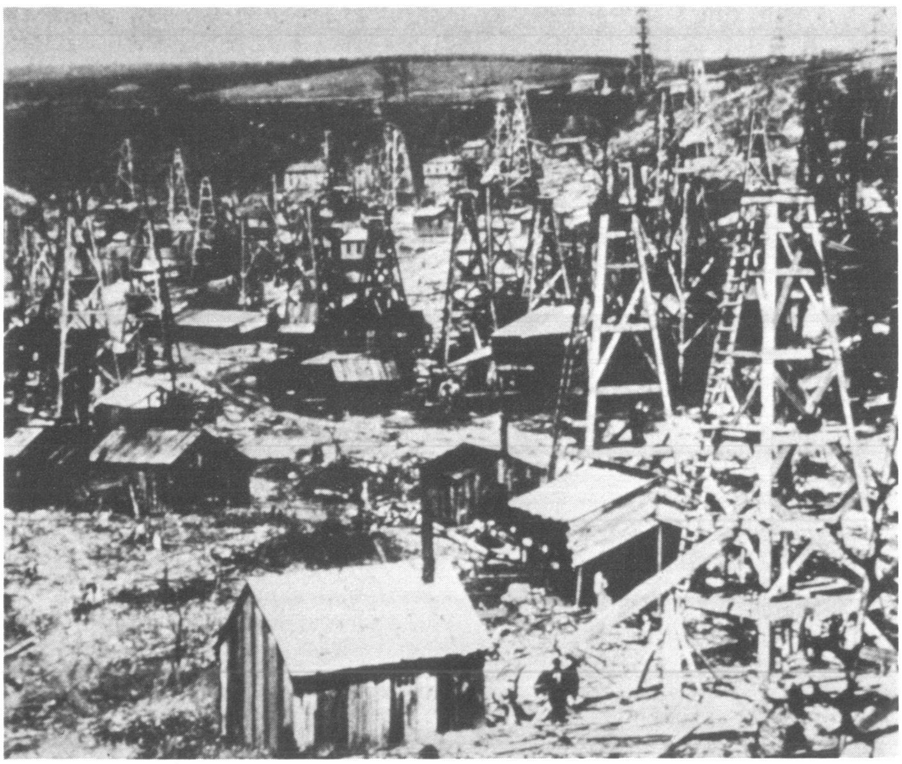
Complex Geology

In many areas, the rock layers of the earth's substructure are not flat and uniform. Forces within the earth have twisted, folded, and faulted layers of rock. Waves reflected from these deformed structures behave in what, from the surface, looks like a nonsensical manner. Simple plotting of the reflection times from the various rock layers yields a very distorted picture of the complex subsurface structure. In those areas where the geologic structure is complex, the geophysicist needs a great deal of additional data processing and plotting to get a useful picture. Although this problem has been challenging geophysicists for years and some significant progress has been made, it is now getting a great deal of special attention. Some of the most promising land areas in the United States remaining to be explored for oil and gas are extremely complex geologically. Each advance in seismic technology that increases the chances of successfully exploring complex structures underneath the earth's surface enhances the odds of finding crude oil or gas.

HISTORY OF OIL EXPLORATION

The Early Years

In 1845 when Robert Mallet was setting off "artificial earthquakes," oil was already in demand. Petroleum, which literally means "rock oil," was being sought and found. For thousands of years petroleum had been used as medicine, a lubricant, and even as a fuel. But the demand for this substance had been met by finding surface seeps where the oil either bubbled to the surface in places such as on the Island of Trinidad, or seeped out of rock formations where stream beds cut into oil-bearing rock formations. In the nineteenth century, however, whale and fish oils were becoming scarce and more expensive. Improvements in refining made petroleum an attractive substitute for the other fuels. By 1857 the demand for petroleum products, such as kerosene, had grown enough to make the idea of drilling for oil economically attractive. Although oil had been found as a by-product of drilling for water, no one had yet drilled specifically for oil. With the market for oil expanding, however, it was only a matter of time until oil wells would make an appearance. The first oil well in the United



The search for oil was on, and a new industry grew almost overnight. This photograph was taken in 1865, on the Benninghoff farm near Oil Creek, Pennsylvania, just 6 years after Drake's successful well was completed in Titusville.

American Petroleum Institute


DRAKE

Coopers or barrelmakers played an important role in the early days of the development of the oil industry. Just 10 years after this photograph was taken in Omaha, Nebraska around 1890, the United States had produced a billion barrels of oil.

American Petroleum Institute

States was dug in Cuba, New York, in 1857. It was a "dry hole." One year later the Seneca Oil Company of Connecticut began drilling a well on the Hibbard farm near Titusville, Pennsylvania. After drilling 69 feet 6 inches, the company struck oil. (Today successful wells are often drilled to depths of 25,000 feet or even more.) The Titusville well was a success. In 1859, its first year of operation, it produced 2,000 barrels of crude oil.

As the demand for oil grew, the search for it became more aggressive. Companies that supplied the growing demand reaped fortunes. The cycle of demand and supply seemed to feed on itself and to grow explosively as new uses for a convenient and cheap source of power were developed. By 1900 the United States had produced a total of 1 billion barrels of oil; the second billion was produced by 1909; and the third billion, by 1913. This phenomenal rate of growth was just the beginning.

On August 20, 1869 Edwin Drake completed the world's first successful oil well near Titusville, Pennsylvania. Drake, in top hat, is pictured in front of the well he drilled for the Seneca Oil Company.

Drake Well Museum

Early petroleum production was achieved without the use of geophysics. Oil was found by locating wells in places where oil had seeped to the surface, by looking for geologic structures on the surface similar to those that had yielded oil in other locations, or by a number of other means, most of which were not very scientific. Explorers had to be willing to gamble, to "wildcat." The stakes were high, but so were the rewards. Many of today's largest oil companies trace their beginnings to this period, but the petroleum industry would not have survived very long if it had continued to depend on comparatively haphazard approaches to exploration. Fortunately, a large number of people in business were willing to risk their financial resources, and scientists were willing to devote their careers to developing better ways to find oil. It was natural that investigators took geological knowledge and tried to marry it to long-known principles of physics to see if there were any promising means of increasing their ability to find oil.

The problem was to discover a method for locating likely oil traps without resorting to the impractical alternative of drilling everywhere—a kind of geological blind-man's bluff. Investigators needed a way to "see" the structure of the rock layers beneath the earth's surface.

By 1921 two Oklahoma physicists were ready to try to look beneath the surface with seismic waves. William Haseman, a physics professor, and John Karcher, one of his former students, began attempting to measure seismic velocities in rock formations in the area around Ponca City, Oklahoma. Their equipment and methodology, although crude by more modern standards, contained the essential elements used in seismic exploration today. Dynamite shots were set off below the surface. The instant of detonation, called the "time break," was marked electronically on film running through a photographic film recorder. The seismic waves were picked up by a microphone detector on the surface, and exact times of their travel plotted electrically on the recording film. Calibrations for the timing were produced by a tuning fork vibrating at precisely 100 times per second. Accurate recording of the times it took the seismic waves to travel from the detonation location, the shotpoint, to the microphone was essential. Haseman and Karcher hoped that they would be able to locate types of structures that are often associated with oil traps. Field tests showed that their equipment was working; their methodology was sound. But the real question was, could their system be used to find oil?

Despite careful design and the years they devoted to testing, results were still disappointing. The project had to be abandoned. Unfortunately, the area their first client had asked them to explore was not amenable to obtaining good data. Haseman and Karcher's equipment was not sophisticated enough to make sense out of the jumbled seismic waves returning to their microphone. The nature of the geology of the area, not a flaw in their equipment or methodology, had defeated their attempts.

In 1923 another major, unsuccessful attempt to locate oil-bearing structures was conducted in the United States by a German company, Seismos, Ltd. This time the seismic exploration for salt domes focused on the Texas and Louisiana Gulf Coast. The luck of the Seismos crew was as bad as that of Haseman and Karcher. It happens that there are no shallow domes in the part of the Gulf Coast they were searching. There are deeper domes in that area, but the equipment in 1923 was not sensitive enough to pick up refractions coming from the greater depths. The following year, however, another Seismos crew contracted by Gulf Production Company obtained useful refractions from a salt dome in the Texas Gulf Coast region—the first seismic discovery of a salt dome that contained oil. The seismic method was beginning to prove its potential.

In 1926 a seismic crew was hired by Gulf Production Company from the Geophysical Research Corporation (GRC), a geophysical contracting company. That crew discovered two salt domes within 3 months. Part of the reason for their success was that they had placed their detectors farther away from their shotpoint than had been done before. They reasoned that by increasing the shotpoint-to-detector distance to as much as 6 miles, they would be able to receive refractions from far deeper in the earth and to collect evidence of deep salt domes. Their results showed that their reasoning was correct. The equipment was the same, but their new methodology made the leap forward possible.

Another of the GRC's innovations was a new amplifier that could tune out undesired "ground roll." Ground roll is a seismic wave that starts with the explosion at the shotpoint, travels along the surface of the ground directly to the detectors, and drowns out the weaker, but desired, reflections coming back from deep in the earth.

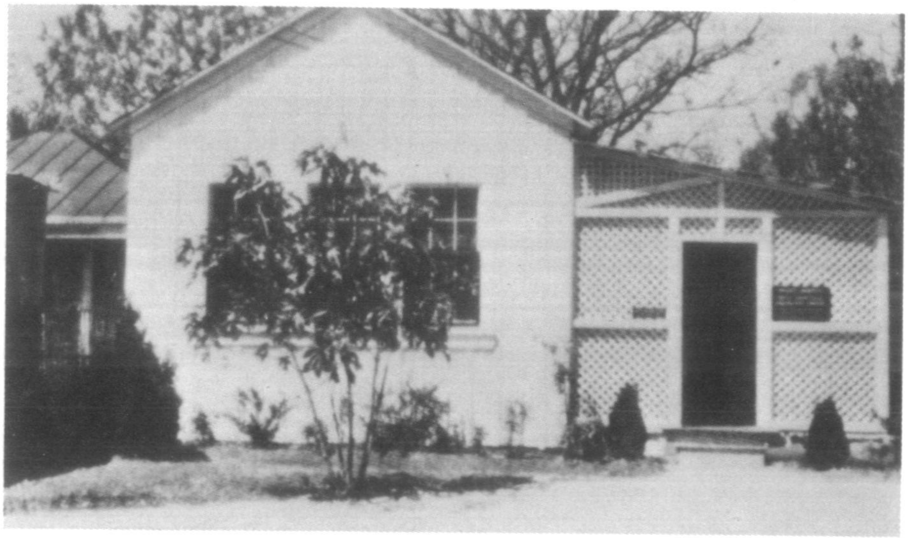
Developments such as these enabled GRC to be the only company during the 1920s to carry out commercial reflection seismograph prospecting. The discoveries they made for their parent company, Amerada Oil, were proving over and over again that the system worked. Because of GRC's outstanding success, Amerada decided to limit the use of the GRC reflection crews exclusively to Amerada itself and to its subsidiary, Rycade Oil Company. Toward the end of the 1920s GRC employed 70 percent of the world's seismic scientists; Amerada's decision, therefore, constituted a serious restriction of the development of seismic exploration.

Everette De Golyer, president of GRC, disagreed with this monopolistic policy of his parent company. In February, 1930 he took an unusual step. De Golyer secretly financed the organization of

Geophysical Service, Incorporated (GSI) to do reflection seismograph prospecting. Because the patents on the seismic equipment he needed were held by Amerada, De Golyer's new group was forced to develop its own designs and prove them to the oil companies. Amerada's successes had made its competitors eager for the new seismic exploration technology. GSI offered them a way into the game. Within 1 month GSI had 10 contracts for as many crews. Geophysical Research Corporation's monopoly was broken. It took several years for them to refine the new equipment, but by 1938 GSI had 34 crews working in North and South America and in the Far East. GSI was on its way to becoming the largest geophysical contracting company in the world.

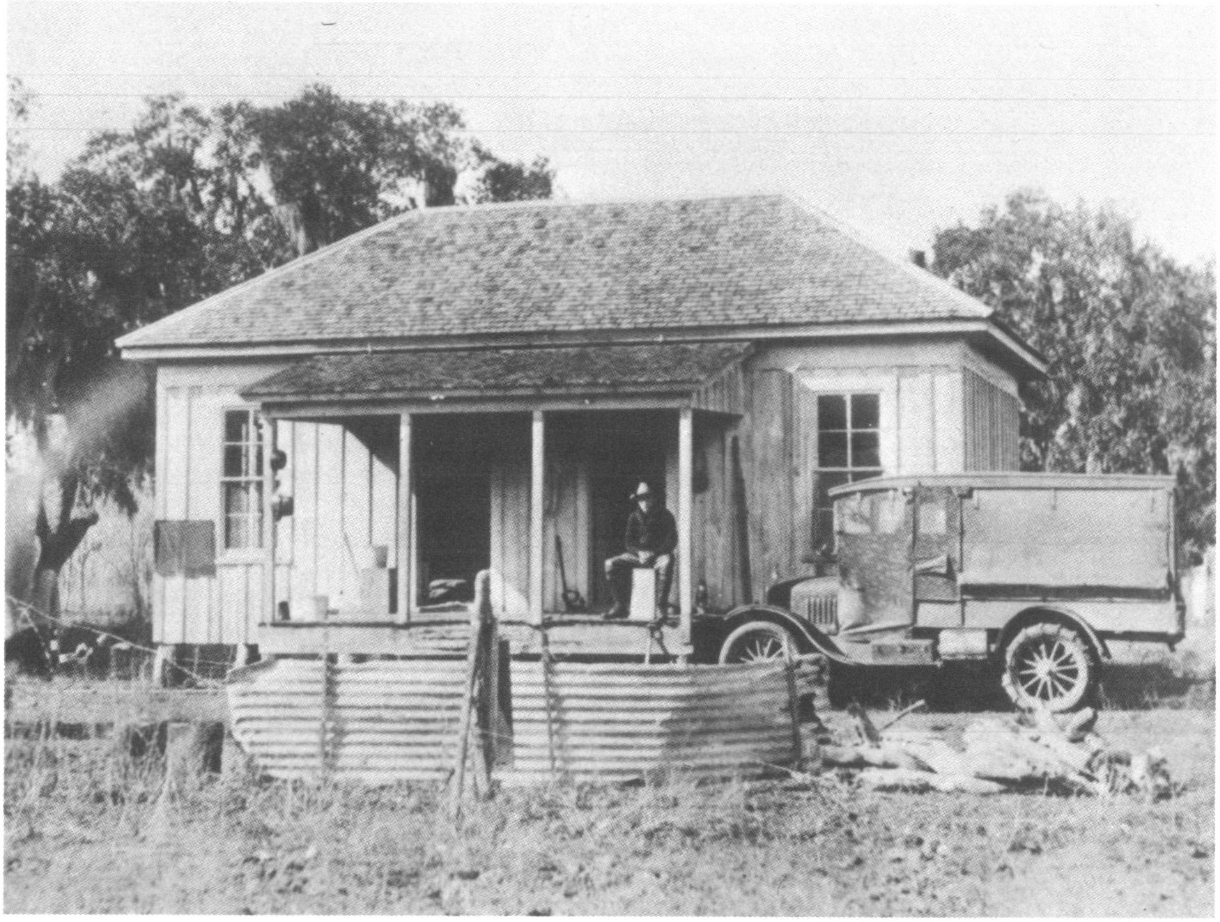
At least 30 new United States seismic contracting companies were formed during the next 10 years or so. Of these, almost all were founded by former GRC and/or GSI employees. Most of these new companies built instruments similar to the new design GSI had developed to get around the GRC patents. However, several of the new contracting companies developed seismic equipment of their own design. One such company was Petty Geophysical Engineering founded by two brothers, Dabney and Scott Petty.

In early 1925 Dabney, a geologist from the Texas Bureau of Economic Geology in Austin, heard about the seismic exploration activities of the Seismos crew working in eastern Texas on timberland owned by the Petty family. Dabney decided that if he and his brother, Scott, could obtain a seismograph and learn how to use it, they could go into the consulting business trying to locate promising sites themselves.



This house is where the Petty brothers developed their prototype of a seismic detector. In Scott Petty's words (from Seismic Reflections*), "Our first 6 months were the hardest for we not only had technical problems but financial problems as well. Our salaries had ceased when we quit our jobs. Our folks helped out by giving us room and board for free and furnishing us a place to work.... The location was ideal for we found that trolley cars, passing over bumps in the tracks two blocks away, set up helpful vibrations in the earth, which were fairly uniform in nature."*

O. Scott Petty and Geosource Inc.



Scott and Dabney Petty set up the headquarters for some of their early testing of seismic instruments in this abandoned farm house in Brazoria County, Texas.

O. Scott Petty and Geosource Inc.

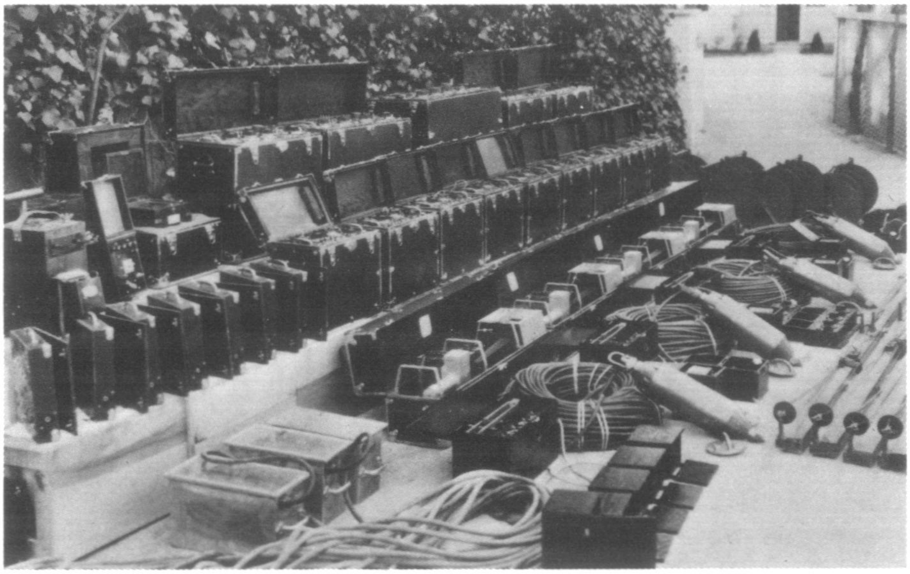
Dabney asked Scott to read everything he could find about the Seismos Company, geophysics, and earthquakes. Coincidentally, Scott had just read about an extremely senstive vacuum tube "ultramicrometer" (a device for measuring very tiny distances) invented by John J. Dowling of the University of Dublin.

Scott ordered one of Dowling's ultramicrometers. Both brothers resigned their jobs and, with $5,000 borrowed from their father, started work at the family home in San Antonio on a very sensitive prototype seismic detector, using the micrometer to measure minute deflections. They first tested the system with the vibrations from trolley cars passing over bumps in the track two blocks away.

After some initial field testing and equipment modifications, they leased acreage partly on and partly off a known salt dome. They spent the winter of 1925 experimenting with refractions, reflections, and dynamite shots of all sizes, set off at a variety of distances to obtain various types of profiles of their dome. They learned how to obtain good salt dome records with as little as 20 pounds of dynamite, rather than the several hundred pounds used on each shot by the Seismos crews, a significant economic advantage.

Their first commercial job was to "shoot" a block of family-owned timberland in the southeastern part of Texas. Because of the results of tests on their salt dome, they believed they could obtain a reflection directly from the salt by setting a detector fairly close to the shotpoint. At the Texas site, they obtained evidence of such a reflection. Their excitement at that achievement faded quickly when they discovered that the evidence was the result of a loose nut that was vibrating on the detector! Despite that disappointment, they subsequently mapped a dome with a relief (the maximum difference in elevation between the highest and lowest points on the top of the dome) of only 150 feet, a notable achievement. The significance of their success was that, since the dome was about 10,000 feet below the surface of the earth and the relief of the dome was only 150 feet, the sensitivity of the equipment, especially the recording devices, was fine enough to distinguish between waves arriving only hundredths of a second apart. (Twenty years later that particular dome was drilled and produced oil.)

For the next 2 years, Dabney and Scott improved their techniques and equipment and had the satisfaction of finding prospects missed by other crews. From 1927 to 1931 the brothers had three refraction crews



A complete set of refraction equipment that was airlifted to Venezuela in 1931 by Petty Geophysical Engineering Company.

O. Scott Petty and Geosource Inc.

in Venezuela; in 1930 the brothers sold a complete set of their refraction instruments to Humble Oil and Refining Company. By 1940 Petty had 17 crews in the United States and 7 in foreign countries; their operation was nearly as large as GSI's.

The 1940s and 1950s

During this period the major oil companies were no longer content to rely completely on the geophysical contractors for crews, equipment, and field expertise. Almost all of these oil companies formed geophysical departments of their own, staffed them with the best talent they could find, and supported active research programs. The oil companies tried to develop all available geophysical techniques into exploration tools. Some of their efforts were quite effective. Humble, applying modern technology to scientific principles discovered more than 100 years before, developed and built the first commercially practical gravity meter. The gravity meter measures the average density of subsurface rocks to determine whether older and more dense rocks have been pushed closer to the surface; it can also determine whether less dense salt has been pushed toward the surface. Using this technique of measuring the earth's gravity at the surface to "see" rock sediments below, that company found 15 prospective sites in Mississippi in 1940.

Many of the geophysics departments of the major oil companies made attempts to develop practical methods of detecting oil and gas deposits directly by running electric current through rock layers. Today, airborne magnetometers, devices for measuring magnetism in rocks, can map the basement structure of an area from an airplane, sometimes giving clues to the structure of overlying sedimentary rock layers. This technique, based on a per-mile cost, is quite inexpensive and can be used to "screen" large areas for geological structures most likely to be favorable for the accumulation of petroleum.

The intimate working knowledge of all aspects of geophysical technology acquired by these and other research groups was of utmost importance in establishing a base for the future of petroleum geology.

World War II had both an immediate and a long-term effect on the development of the seismic method. The national effort to develop new weaponry for use in the land, sea, and air battles of World War II was unprecedented. Geophysical techniques that had been developed to find oil were adapted, refined, and improved to contribute to ways of solving war-related problems. Technology and methodology helped develop better ways of locating enemy submarines (sonar) and artillery (echo-ranging). Many of the geophysical laboratories found they had a great deal to offer to the national defense effort and devoted at least a portion of their facilities and personnel to the war effort and to military projects. Although this deflection of personnel and resources considerably slowed the development of seismic technology until a few years after World War II, the technological advances stimulated by war in fields which related only peripherally to seismic exploration were turned to great advantage by geophysicists when the war was over.

Wartime advances in the field of electronics led to the postwar development of smaller, lighter, and more sensitive geophones. The improved geophones could be adapted to new uses. Before the war, for instance, it had been standard practice to use a single geophone for each channel, usually recording four channels of data. A single geophone will pick up and record confusing seismic waves on the surface as well as the unwanted reflections from depth. New developments made it possible to put multiple geophones together on a single trace, or channel, thereby providing a way of cancelling surface waves arriving at the different geophones at different times and of reinforcing reflections all arriving at the same time. This technique resulted in seismic recordings with a significantly improved signal-to-noise ratio. Some of the "snow" was gone; the postwar geophysicist was getting a little clearer picture of the earth's substructure.

Advances in electronics and in instrumentation stimulated by the war also had an effect on recordings. Before the war, a four-trace record of seismic waves had been fairly standard. By 1950 recordings using 24 channels were common. More information and more useful seismic clues, therefore, were at the disposal of the oil prospector using the seismic method.

In addition to improvements in the detector at one end of the system and in the recorder at the other, there were advances in the design of the amplifier which connected them. A seismic wave reaching a detector is transformed by the detector into a weak electric signal. Before it can be recorded, that signal must be made stronger or amplified, a process not unlike turning up the volume on a radio. In fact, advances in radio and television equipment led to the development of the new generation of amplifiers. Amplifiers were placed on each channel to boost the output of the geophone groups. Each one of those 24 individual trace amplifiers was provided with easily adjustable volume controls and with special high-frequency and low-frequency filters to help cut out various types of unwanted noise.

Another "crossover" advance from radio and television technology came with the concept of automatic gain control (AGC). The AGC automatically adjusts the volume of a signal to a desired level, even if the strength of that signal varies. Today AGC is best known as a standard feature on many home tape recorders; it was also an important advance in the seismic method. In seismic recording AGC is used to adjust the volume on the recorder as the progressively weaker reflections coming from increasingly deeper levels arrive at the detectors on the surface. Such faint echoes, once lost to the geophysicist, are now available for interpretation.

Using this new, postwar equipment, geophysicists were able to investigate areas of far more complex geology than had been possible in the past. New areas were opened up for the search for oil. While the new methods did require a lot of slow and tedious work interpreting the growing mound of data recorded from each seismic shot, they were a definite improvement over the prewar methods.

By the early 1950s the growing demand for oil had seismic crews searching underwater for oil. A considerable amount of the seismic work was being done in shallow-water areas such as the swamps and bayous of Louisiana. Instead of being put on trucks, the recording equipment was put on boats or barges, or on specially constructed "swamp buggies." The seismic method was proving to be not only accurate, but also adaptable to the widened search for oil and gas.

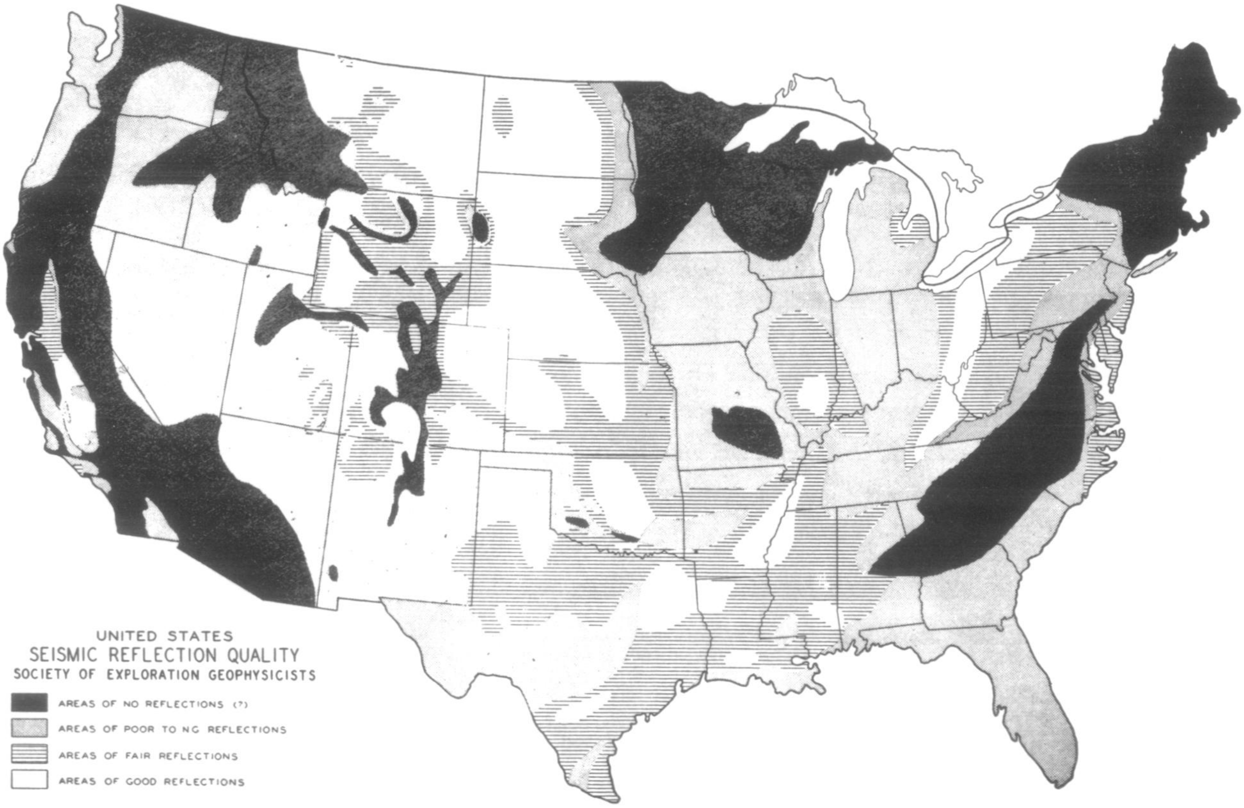
Although seismic shooting in open-water areas such as the Gulf of Mexico was underway during this period, offshore exploration was not yet a large percentage of the overall seismic exploration effort. (The first separate cataloging of marine seismic efforts in the Geophysical Activity Report of the Society of Exploration Geophysics was not until 1953.) Strangely enough, because of World War II, more was known about the physical processes involved in underwater explosions than was known about those that occur in a shothole as part of land exploration. The base of knowledge gained from wartime research was to be an important factor in the success of the seismic method in marine exploration, and marine exploration was the seismic wave of the future. Many of the major oil finds since the late 1950s have been offshore in open water. Such finds, including the opening of the vast North Sea fields, were made possible by information gained from marine seismic exploration. Without those major undersea fields, the world supply of crude oil would have been considerably lower, resulting in increased economic hardships for all nations depending on imported oil.

The description of the state-of-the-art of seismic prospecting in the 1950s may give the impression that the technology available was equal to the exploration task at hand. This might have been true had it not been for two extremely important factors. First, large areas of the United States and the world could not be explored with commercial success because of very bad noise problems, very complex geology, or very elusive traps. Second, petroleum prospects in the good shooting areas were becoming increasingly difficult to find.

In the 1951 seismic-reflection quality map of the United States shown below the dimensions of the first problem are quite clear. In the potentially productive areas of the country, large sections yielded either "poor to no-good" reflections or only "fair" reflections. Twenty-five percent of the area of the United States has little chance of producing oil because of its geologic history. Only 17 percent of the prospective area was judged to give "good" reflections, and only about 28 percent, "fair."

This map shows the seismic reflection quality of the United States in 1951.

Assembly by Paul Lyons



The challenge was plain: first to develop improved techniques that would yield better results in the more than half of the area in the United States where oil might be found but where the complex geologic substructure made seismic exploration impractical; second, to improve techniques for finding petroleum in the good shooting areas where petroleum prospects were becoming more difficult to find. This decline in petroleum prospects can be seen by looking at the production results of an average seismic crew. In 1934, the golden era of salt dome exploration, a single geophysical crew could expect to find about 14 million barrels of oil in a year. But almost 25 years later, despite all of the significant improvements in the methodology and technology of seismic exploration for oil, that same crew could expect to find less than 2 million barrels. And they would spend more time, effort, and money to find it. While the demand for oil was growing, most of the "easy oil" had already been found.

Clearly, there was a great need for a new or an improved technology that would make it possible to prospect more successfully for oil in that large area of poor reflection quality. Furthermore, data processing techniques were needed which would enable the interpreter to make use of *all* of the valid reflection information on the seismic recordings—not just the small portion that the geophysicists had time to pick, correct, and plot by crude manual methods using a desk calculator. These improvements were on the way.

Modern Developments in Seismic Exploration

It has become common practice to record musical groups, rock stars, and symphony orchestras alike on multiple sound tracks, with the output of several microphones being recorded separately but synchronously. The advantage of this method is that after the recording session, the engineer and the producer can adjust the relative volumes, the balance of individual instruments, singers, and orchestra sections on the various tracks. They can experiment with filters on individual sound tracks until they are satisfied that they have the best possible sound. The same developments that made this technique possible for the recording industry also made possible tremendous improvements in seismic exploration techniques.

For geophysicists to be able to handle the problems of noise and resolution more effectively, they had to be able to make multiple-track field recordings that could be played over and over again for analysis and could ultimately be combined in the best possible manner for final recording and analysis. Magnetic recording on tape had become a practical reality during World War II. The technology continued to improve in the years following the war, and geophysicists began experimenting with it. They were looking for a way of recording information from their shooting in the field that would be a significant improvement over the limited methods then in use.

One system, developed by Field Research Laboratories of Magnolia Petroleum (now Mobil Oil Company), allowed geophysical crews to record synchronously 13 separate tracks or channels over a wide range of frequencies. By experimenting with many playbacks of the recording, the geophysicist was then able to choose the best filter setting or settings and tune into the precise frequencies that offered the best results in that particular circumstance. The results of seismic exploration done with this new generation of the record/playback system successfully demonstrated the advantages of the new generation of exploration equipment.

As a result, the development and the manufacture of equipment for the magnetic recording of seismic events began to mushroom. Geophysical Service, Incorporated, a very large geophysical consulting firm, had become heavily involved in the design and manufacture of new instrumentation during the Korean War. That involvement prompted the founding of their subsidiary, Texas Instruments (TI), in 1952. The geophysical company eventually was able to take advantage of TI's pioneering work in semiconductors by developing new magnetic recording equipment designed specifically for the seismic exploration for oil. By 1954 TI was ready with a magnetic-disc field-recording system and an advanced amplifier. It was the first in a series of new systems and set a new standard for the industry.

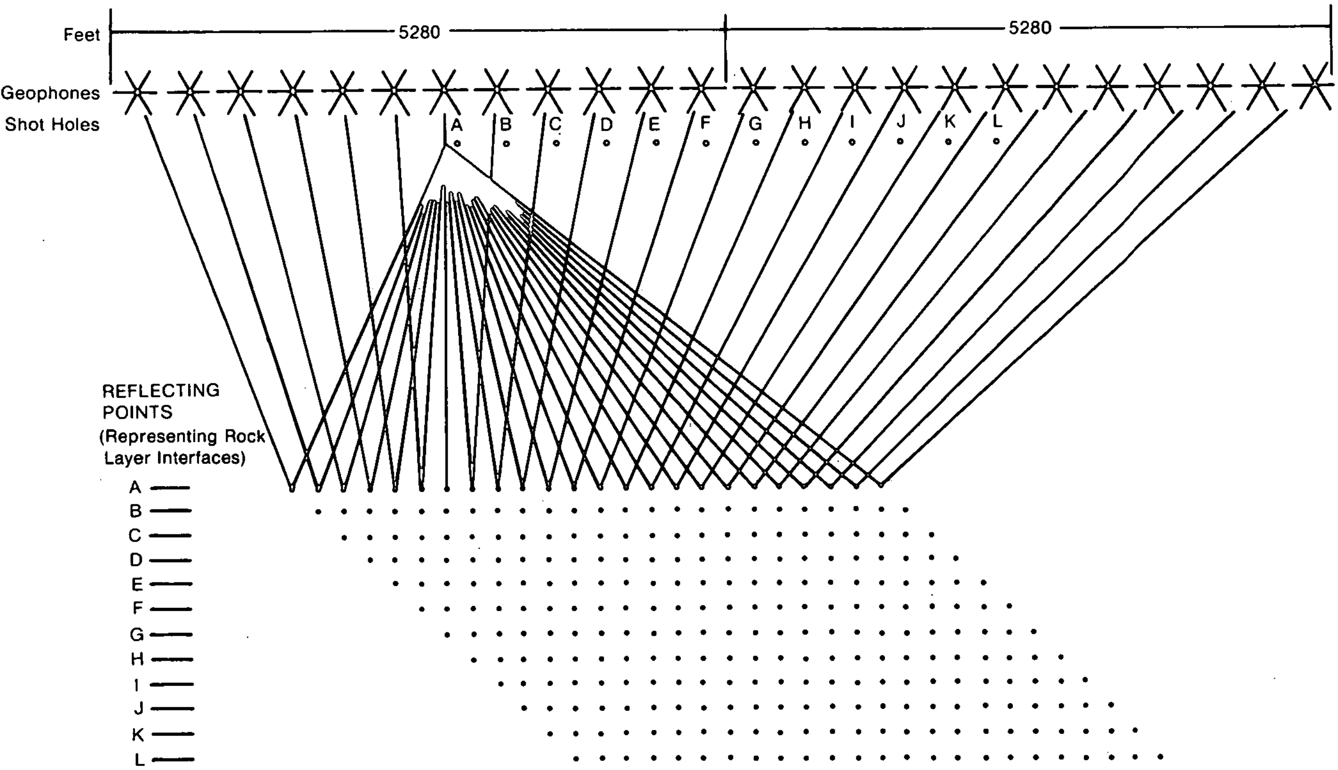
Other companies rushed into development also. In 1955 more than 180 magnetic recording units were either in use or on order from the 7 manufacturers producing this equipment. These units were distributed among 20 percent of all the seismic crews in the world. The magnetic recording revolution was well underway.

One very effective use of this new technology was to come from a new method of shooting made possible by this recently developed equipment. It is called the common-depth-point (CDP) or the common-reflecting-point (CRP) shooting and processing method.

The best way to understand the common-depth-point method of shooting is to look at an example of it in use. In the figure on Page 65, 24 geophones are spread out in a line 2 miles long on the surface of the area to be investigated. The diagram shows raypaths of the soundwaves generated by the explosion at shothole "A" (below the weathered layer) as they travel to the first line of reflecting points, representing a rock layer interface, and are reflected to the geophones on the surface.

The next shot takes place at shothole "B." In preparation for that shot, the seismic crew will remove the geophone array at the left end and add another at the right end. They will then explode shothole "B." The results of this shot are illustrated by the second row of reflecting points at the bottom of the diagram. This leapfrog process is repeated over and over to shothole "L." By the time all 12 shotholes have been fired, each reflection point will have been covered by 12 different shots

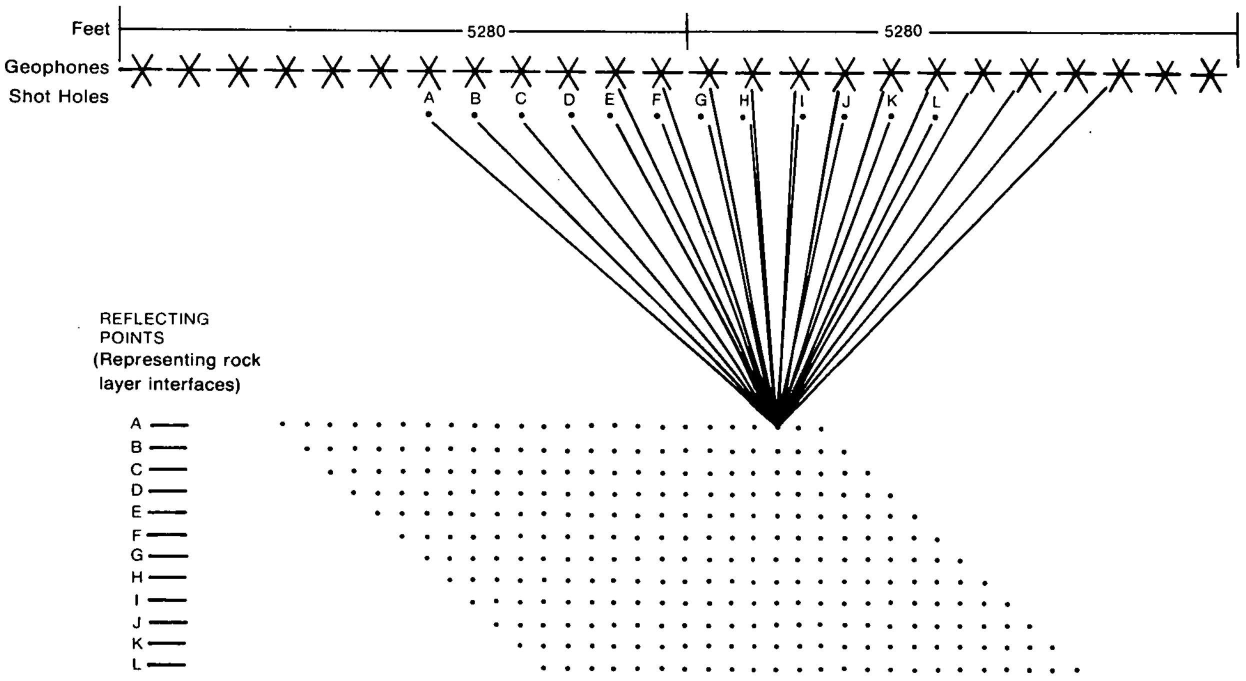
This diagram illustrates a 24-trace roll-along spread with reflecting points for 12-fold addition. The ray-paths of the sound waves generated by an explosion at shot hole "A" are shown.



from 12 different angles as shown in the figure below. This type of shooting provides "1,200 percent" or "12-fold" coverage of the subsurface. If shots were spaced twice as far apart, the coverage would be one half, "6-fold," or "600 percent." The amount of coverage required is dependent upon how much the signal-to-noise ratio needs to be increased. And this in turn is dependent on the specific geology of the area being explored.

As mentioned earlier, seismic echoes from each shot travel to the 24 geophone arrays. The output of each array is recorded separately on a 24-trace magnetic recording tape for playback and interpretation later by the geologist. Although the information recorded in CDP shooting requires a great deal of sophisticated data processing to make the system work, this method has become one of the most valuable seismic exploration tools in use today.

This diagram illustrates a 24-trace roll-along spread with reflecting points for 12-fold addition, showing ray paths for a typical depth point at the level of rock layer interface "A."



The original concept of CDP or CRP shooting, recording, and analysis originated with Harry Mayne of the Petty Geophysical Engineering Company. Mayne conceived the idea when he faced some serious problems on one particular job in 1950. Experience told him that the usual shooting techniques would not give the results needed. The geology of the area was complex and required a large number of shots. These shots, however, would override the desired seismic reflections with noise. Mayne theorized that the arrangement, which today is called CDP shooting, could provide the information he needed. In 1950, however, there were no practical ways of making a useful recording. Even if he could have made the recordings, the processing of the huge amounts of data generated by the multiple shots and large numbers of geophone arrays would have been a gigantic task, too great even to be considered using existing technology. These problems were enough to make CDP an impossibility—for a while. Simultaneously parallel developments in other fields were about to provide answers to the problems of CDP shooting.

The first such development was in magnetic tape. By 1955 tape technology had advanced to the point where it was practical to make field recordings with enough channels to record the output of all the geophone array necessary for the CDP system. When Mayne realized that one obstacle had been overcome, he turned his attention to the problem of processing massive amounts of data, the remaining stumbling block to implementation of CDP. He made the rounds of all the geophysical instrument manufacturers trying to find a company interested in his problem. Mayne recalls the way the equipment was developed. His description is an insight into the complexities of applying even very promising scientific and technical knowledge in "the real world."

> All of the geophysical instrument manufacturers were approached with an outline of the requirements for processing equipment which could perform the necessary correcting, transcribing and summing operations. With the exception of one company, the Texas Division of Brush Electronics, no one could understand the need for transcribing the corrected data to another tape and refused to consider making such a machine.
>
> After considerable negotiation,Brush prepared preliminary specifications and submitted a price quotation on a suitable machine, and Petty entered into a development contract in late 1955. Brush was a newcomer in the geophysical instrument business, and not fully aware of the high standards required by the industry. I [Mayne] suspect that their education was a rather painful engineering and financial process. In any case, after

two years elapsed time, considerable engineering trauma, and a price escalation to 2½ times the original price quotation, they developed quite a competent machine.

Unfortunately, the geophysical business had suffered a severe recession in the interim, and Petty as a relatively small contractor, was dubious about the business wisdom of risking $250,000 (the final purchase price) on an untried instrument designed to perform a process which had not yet been accepted by the industry.

It was a classic case of the chicken and the egg. We could not demonstrate effectiveness of the CRP method without the machine, and we didn't want the machine unless the method proved to be economically viable. All of this against the backdrop of a widespread "it will be too expensive even if it works" expression of opinion from a large segment of the industry.

Consequently, it was with some relief that we received an inquiry from a major oil company requesting us to relinquish our order for the machine to them. Negotiations were completed between the three parties at the SEG [Society of Exploration Geophysicists] annual meeting in Dallas in 1957, and they assumed our order and subsequently took delivery on the first machine of its type. It should be gratifying to all concerned that this particular machine continued to perform creditably until displaced by digital processing in the mid-sixties.

General industry acceptance was still some time away, and the first clients who supported this early work were the Texas Gulf Producing, Pure Oil and El Paso Natural Gas Companies. Through the interest of these companies we were able to develop a considerable body of experience in a variety of areas, and confirm our hopes that the method offered significant potential for data quality enhancement in most problem areas.

This information was gradually percolating through the industry, and a great deal of interest was apparent by late 1960 at the SEG annual meeting in Galveston. As evidenced by licensing activity and the proliferation of specialized data processing equipment designed for the method, interest increased rapidly, from 1960 through 1963, and the last major remaining skeptics had finally accepted the method by the end of the 1960's.

Mayne goes on to describe the chain reaction set in motion by CDP use. This reaction serves as an example of how a new technology, developed for one specific purpose, can lead to a wide variety of changes and improvements affecting an entire field. He writes of this time:

Paralleling a part of this period, another technical development was gaining momentum, namely the digi-

> tal recording and processing revolution. This was also a fortuitous circumstance because the CRP method and digital processing were mutually synergistic.
>
> The widespread use of the CRP method provided the masses of data required to make digital processing attractive....
>
> One other synergism between apparently unrelated technical developments can be cited between the CRP method and the development of non-dynamite sources, both land and marine. The development of these inexpensive and effective new sources undoubtedly impacted acceptance of the CRP method by improving its cost-effectiveness.

Mayne was referring to the fact that using dynamite as the seismic source for CDP shooting requires a lot of shotholes. Drilling and dynamite already were becoming a very expensive part of seismic exploration. Surprisingly CDP, although an expensive process itself, was beginning to reverse this trend and to make possible financial savings, at least in some cases. In other cases, the better energy penetration that geophysicists can get from exploding dynamite in shotholes drilled below the weathered zone and the additional information about the thickness and velocity of the weathered layer obtained during the drilling still outweigh the higher cost. Even in the United States where CDP is heavily used, dynamite still remains the most commonly used seismic source on land.

However, largely because of the effectiveness of CDP, an earth shaking method developed by Continental Oil Company is running a close second. Called Vibroseis®, the method uses trucks on which large vibrators are mounted; the vibrators are used as a seismic wave source. At what would be the shothole locations, Vibroseis® creates seismic waves by vibrating the surface of the ground, usually for a few seconds. It then is moved to the next "shotpoint" and the process repeated. There is a disadvantage to Vibroseis®. Vibration sources require an even more complicated data processing method to determine the exact travel time of a reflection of the vibration than the method needed to measure the reflection from the very short impulse generated by a dynamite shot. There are circumstances, however, in which the method's inherent advantages outweigh this consideration. For instance, vibration can be used to search for oil in urban areas such as Los Angeles, where the use of dynamite is prohibited; in addition, the required complex data processing method helps overcome certain types of noise.

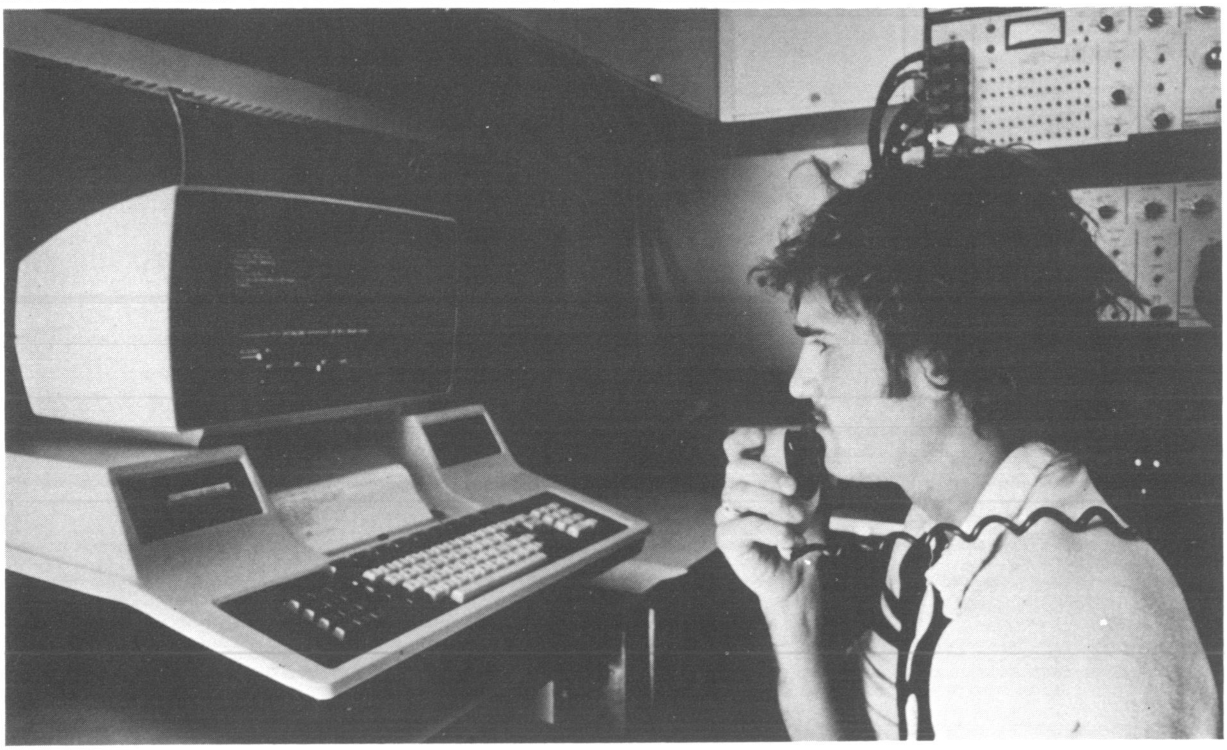
In addition to Vibroseis®, there are other ingenious nondynamite techniques in use. In one system, a seismic wave is created when a gas

Vibroseis® is a technique of exploring for oil and gas which is particularly useful in areas where the use of explosives is not practical. The photographs on these two pages illustrate this method. Several large trucks proceed in tandem along a previously established path. At intervals the trucks stop and a computer synchronizes the lowering of a large steel foot from underneath each truck. The foot lifts the huge vehicle off of the ground, and then the foot vibrates in unison with those on the other trucks, sending shock waves into the earth. These vibrations are detected by geophones placed at intervals across the surface of the land. The vibrations are transmitted to a computerized recording truck located nearby. The trucks are then lowered to the ground, they advance a short distance in formation, and the sequence is repeated.

Marathon Oil Company

is exploded inside a container on the surface of the ground. In another system, a heavy weight (approximately 3 tons) is dropped to the ground. The purpose of both systems is the same as that of the dynamite explosions—to deliver a strong but sharp impulse to the earth. As with Vibroseis®, both of these systems offer cost advantages in getting the kind of coverage required by CDP shooting.

For underwater seismic surveying the variety of energy-producing devices used to generate seismic waves is even greater. Originally, dynamite was the seismic wave-source underwater, but in the early 1950s it was replaced by nitrocarbonitrate (NCN) because the dynamite explosions were killing too many fish. However, even if dynamite had not affected aquatic life at all, the growing popularity of



the underwater CDP method for marine exploration gave the new nondynamite sources a strong economic advantage.

CDP profiling in water, as on land, requires that the energy source be capable of being fired at regular and frequent intervals. For marine exploration the firing takes place at the end of a towline behind the shooting vessel. The explosions must be uniform and powerful enough to generate seismic pulses that will travel through the water and penetrate through the ocean bottom to the rock layers beneath.

The most widely used of all the nonexplosive sources is the air gun. Although air guns had been used originally in academic studies of the ocean subbottom and to research sound transmission through water, geophysicists saw in the air gun an answer to their problem of how to generate seismic waves under water.

To create these waves, air pumps build up extremely high pressures (as much as 2,000 pounds per square inch) in the submerged gun. The instant this pressurized air is allowed to rush through the large holes opened in the sides of the air gun, it strikes the surrounding water with enough force to simulate a small underwater explosion. The pulse from this "explosion" spreads out from the source. The descending portion hits the bottom and generates the seismic wave that will be reflected by underlying rock layers back to detectors (in this case, the pressure-measuring devices are called "hydrophones") being towed through the water above.

One important sidelight to marine searches for oil is the difficulty of exploring areas where the ocean bottom is composed of a hard, consolidated material, rather than the more usual soft mud. In such a case, a large portion of the energy from the descending water pulse is reflected at the ocean bottom, leaving only a weak signal to penetrate the subsurface. At the same time, the reflected pulse bouncing back up from the hard bottom creates high-energy noise in the water. This in turn creates a very difficult signal-to-noise ratio problem which has not yet been satisfactorily solved even by today's highly sophisticated data-processing techniques. Certainly, much more attention will be given to this problem as the search for oil in the ocean increases.

As mentioned earlier, common-depth-point shooting and magnetic recording are providing a great deal more data than did the earlier methods of seismic exploration. From the moment the scientific world became aware of the emergence of computers, geophysicists began thinking up ways to use them to handle seismic data.

In 1950 at the Massachusetts Institute of Technology, a professor of mathematics, G. P. Wadsworth, and a graduate student, Enders A.

Robinson, began to study ways of using digital computers to enhance the interpretation of seismographic records. The results of their early work were promising. Within 3 years, with the sponsorship of most of the major oil companies and some of the geophysical contractors, the study expanded. More staff and graduate students were attracted to work on the project. In comparison to today's methods, the MIT work was primitive. Paper field-recordings had to be hand-digitized for analysis on MIT's "Whirlwind" vacuum-tube computer. The task was not only time-consuming, it was also tedious.

The method of analysis these researchers were using depended heavily on the work of a number of theoretical mathematicians and statisticians. Although that earlier work had provided a strong theoretical base for geophysical application, there was still much left to be done. New theories and techniques applicable specifically to seismic waves had to be developed. The work went on for 5 years.

Luckily, as the task continued, computers themselves were being improved. Although they were slow and small in capacity by today's standards, these new machines permitted simpler types of seismic analyses to be performed on a practical basis, replacing some of the more tedious calculations with earlier computer programs. There was a fairly gradual transition through several stages from the use of analog magnetic tape recording and minimal processing, to analog recording with digital processing, to digital recording with digital processing.

By the time the second generation of general-purpose digital computers became available in the mid-1960s, most of the oil companies had switched to digital field-recordings. Some of them started to do data processing themselves—both for research purposes and to try to improve upon the geophysical contractors' processing results. Core memories of 32,000 words or more were available by then, along with multiple times in the microsecond range, so that the processing rapidly became even more sophisticated. Although costs were still high, they were gradually being reduced by the development of new techniques for handling data and by continuing improvements in computer technology.

About this same time, a promising part of the North Sea was being explored for sites likely to produce oil. The processing of data from this area was difficult. The ocean subbottom in this region, with its complex and highly variable geology, stretched all of the state-of-the-art digital processing techniques to the limit of their capabilities.

The development of velocity analysis methods greatly helped to resolve this impasse. Velocity analysis is the calculation of the speed with which a seismic wave travels through a rock formation. As mentioned earlier, velocity is an extremely important factor in seismic exploration not only because it can be used to determine the depth from which the reflections return but also because it provides valuable clues to the types of rocks present. In the mid-1930s it became routine to "shoot" wells for velocity. These velocity surveys were done by lowering a special geophone down the well hole and measuring the exact travel times to different depths by shooting at the surface. While this technique gave accurate velocity information at the well, the data were not necessarily applicable elsewhere.

Not until the advent of CDP shooting, digital processing, and the development of velocity analysis programs did it become possible to routinely determine useful velocity information from regular data throughout the entire survey area. The combination of the seismic method and digital processing made the accurate surveying of parts of the bottom of the North Sea possible and led directly to one of the largest oil discoveries of recent years. The effects of that one discovery on the economies of the participating nations and on the world supply of oil will ultimately affect almost every person in the world.

In recent years the ability to process seismic recordings has continued to improve, and industry reliance on computers has continued to grow. Today the United States petroleum industry uses more digital computer power than any other private industry in the world. Within the petroleum industry seismic data processing is the largest user of that computer capacity.

As oil has become increasingly harder to locate, the pressure for techniques that will locate more subtle configurations that may contain trapped oil has increased. New instrumentation and new computer programming have led to the development of new "high-resolution" techniques. These methods require even more "number crunching" to extract all possible information from seismic field-recordings and to see the earth's substructure more clearly. The results are impressive. One of these new high-resolution systems has demonstrated that it can "picture" a layer as thin as 25 feet that is almost a mile below the surface. Such techniques are expected to prove extremely valuable in finding oil.

A number of problems in the use of the seismic method have been at least partially overcome. Although much research remains to be done, the seismic system has become a major tool for the geophysicist in the difficult task of finding geological conditions capable of entrapping oil and gas.

OTHER USES OF THE SEISMIC METHOD

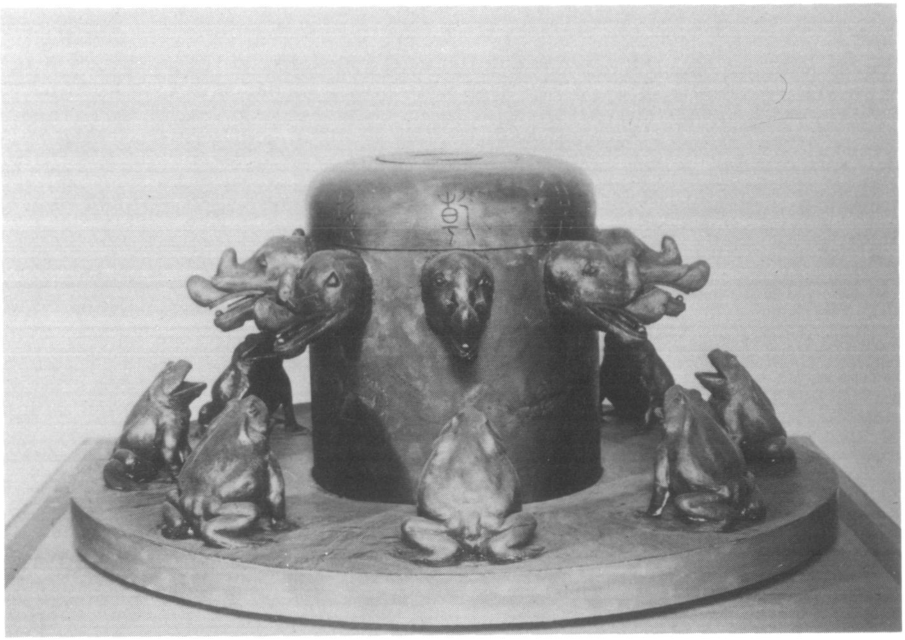
While its role in the search for oil and gas is probably the most widely publicized, the seismic system has made significant contributions to other areas as well.

In the mining industry, the primary traditional applications of the seismic method have been to map the depth of rock layers down to bedrock or to explore for ancient channel deposits that may contain uranium. More recently, high-resolution techniques for seismic exploration have been modified to map coal seams. The spread of the seismic method into the mining field has been somewhat limited by a lack of appreciation for seismic techniques on the part of mining geologists. But acceptance of the seismic method is growing as it continues to prove itself as a cost-effective means of collecting valuable geological information. The increasing value of minerals has made the seismic method (a very expensive way to locate minerals) more economically feasible.

The seismic method is also being used to study potentially important geothermal power sites. The structural information collected about geothermal power sources deep inside the earth will greatly facilitate the development of these energy resources.

Some of the refinements in the technology used in the search for oil have also been adapted to study earthquakes, just as some principles of studying earthquakes have helped geophysicists locate hydrocarbons. The scientists who study earthquakes and their behavior, causes, and even their prediction are working with techniques that are familiar to the petroleum geophysicist. There are also similarities in the equipment that is used. The instruments rely on the same scientific principles; the major difference is in the source of the "bang." In seismic exploration the vibrations are set off when and where the investigator wishes. With earthquakes, movements of layers of the earth itself generate the vibrations.

The earliest known "seismograph" was invented in China about 2,000 years ago. Choko, the inventor, designed a sculpture; it was a circle of dragon heads, each with a copper ball in its mouth, surrounded by a circle of frogs. When the earth shook, the copper ball fell from a dragon's mouth into the waiting mouth of a frog (page 76). As quaint as this arrangement may seem, the Chinese reported that this device detected an earthquake about 250 miles away. Scientists who study



This model of Choko's seismograph was made in 1956 by referring to an old Chinese drawing.

National Science Museum in Tokyo, Japan

earthquakes today are interested in knowing not only that an earthquake took place but also how strong it was, how long it lasted, the degree of horizontal and vertical movements of the earth's crust, where in the earth's crust it was generated, the kinds of motions that took place, and when and where the next one might occur. The answers to such questions are helping to reduce loss of life and damage to property from one of the most destructive forces in nature.

Americans tend to think of earthquakes as a problem of some very specific parts of the country, especially of California. They forget that cities such as Boston, St. Louis, and Charleston have also been shaken by severe tremors. Because earthquakes can occur in a number of locations, studies using the principles of the seismic method have been useful in choosing locations for structures for which the prevention of damage from earthquakes is particularly critical. Such structures include nuclear plants, dams, public buildings, skyscrapers, and bridges.

Vibration detection devices similar in principle to those used in oil exploration have been modified to record how different building materials and various architectural designs react to simulated earthquakes. Scientists are using the information from these tests to engineer structures that will withstand the tremors of quakes, safeguarding the lives of people who live and work in those sections of the world threatened by the awesome power released when the earth's crust shifts and readjusts to the constantly changing pressures originating far beneath the surface. In California, this knowledge about the behavior of buildings and construction materials has already modified architectural and engineering practices significantly.

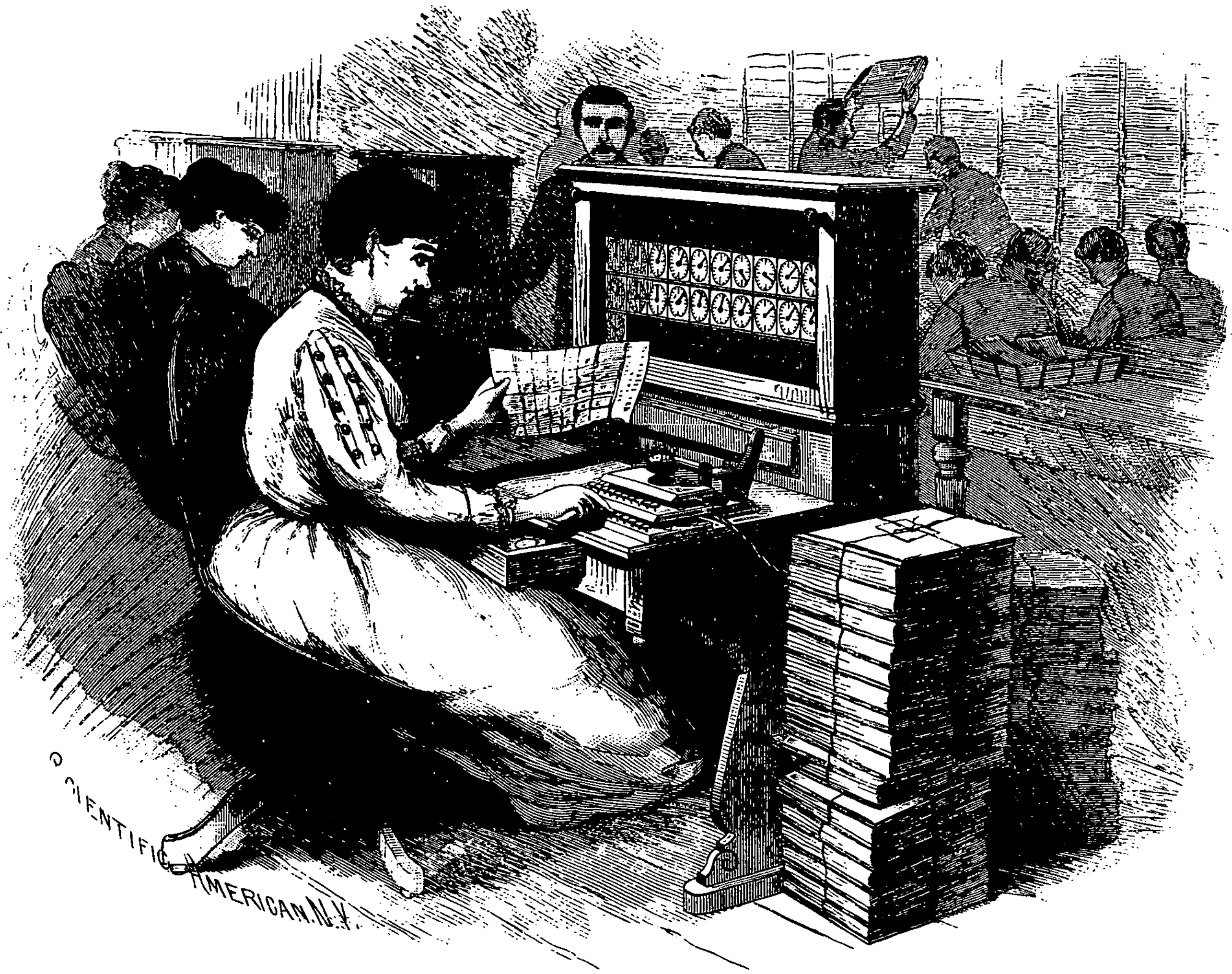
The seismic method is also helping in the search for answers to basic questions concerning the geological history of the earth, the earth's structure, and the forces that are responsible for movement of the continents. A major long-term research project has already increased understanding about some of these basic phenomena. Groups with an economic stake in knowledge of subsurface geology, aware of the potential for significant practical applications in the near future, are watching this project with intense interest.

This remarkable project is known by its acronym COCORP (Consortium for Continental Reflection Profiling). COCORP's purpose is to investigate the make-up of the earth's lower crust and upper mantle throughout the United States—a heroic undertaking. Using seismic equipment, scientists are "seeing" more than 35 miles straight down into the earth. As is usually the case in basic research, no one is able to predict fully either the results or the effects of the research. After only a few years of investigation, however, the results are impressive. COCORP's research has provided some startling and important answers to questions that have puzzled geologists for years.

In its investigation of the Appalachian Mountain region, for instance, what the COCORP researchers learned has staggered the field of geology. They have learned that the substructure of the Piedmont region, the geologic region east of the Appalachian Mountains, has been seriously misunderstood. The cause of the misunderstanding is a mass of rock that has been thrust westward at least 100 miles over thick ancient oceanic sediments. This overriding rock layer, 20,000 to 45,000 feet thick, had been thought to be basement rock, under which there could be no sedimentary deposits. The existence of this "thrust fault" (basement rock that has been shoved over sedimentary strata as a result of an earth movement) is a very important piece of basic information for people who are trying to understand the earth's geologic history. The knowledge is revolutionizing concepts regarding the formation and development of such mountain belts as the Appalachians and the Ouachitas. Geologists and geophysicists are excited by the possibility that such thrust faults may be more common than had been believed. It is important that COCORP expand to test other regions with new hypotheses developed during the Appalachian work.

Scientists are not the only people fascinated by the results of the work in the Piedmont region. The sedimentary strata under the overthrust rocks turn out to be very much like rock layers that often contain petroleum of one form or another. In addition, the dense crystalline rock thrust over them makes an almost ideal seal for oil and gas which might be stored in porous sedimentary rock below. In fact, as a result of COCORP, several oil companies have sent seismic crews to the Piedmont area in the hopes of finding oil and gas. Only exploration wells can show whether there is commercial oil or natural gas available within those very deep sedimentary strata of the Piedmont. The findings of COCORP have suggested that portions of the United States previously considered barren may contain badly needed oil and gas.

Whether searching for oil at depths of 25,000 feet or looking for clues to the geologic history of the earth at depths of many miles, the scientific principles involved are similar. On a COCORP expedition there may be university professors and experienced oil explorers working side by side, each able to make an important contribution of knowledge and experience toward the solution of the problems at hand. The university community and the oil industry have learned that the principles of seismic exploration provide a versatile and accurate tool for seeing beneath the surface of the earth. The seismic method owes much to basic science, the roots of which penetrate the history of science from the development of the silicon chip back to Newton, Galileo, and Da Vinci. It is today successfully and increasingly repaying that debt.



Herman Hollerith suggested that cards with holes punched in them could be adapted to the tabulation process. Such cards were used in the machines he developed for recording the United States census of 1890.

Survey Research and Opinion Polls

In 1888 James Bryce, a British lawyer and Member of Parliament, noted a "new force" in the world, "conspicuous only since governments began to be popular." That force was public opinion. Lord Bryce believed that whether a country is despotically governed or free, "both are generally ruled by opinion." But in free countries people *feel* their supremacy, treating their rulers as their agents, while in despotic countries the habit of submission prevails. Bryce felt that a stage in the evolution of opinion, from unconscious and passive to conscious and active, could be reached "if the will of the majority of the citizens were to become ascertainable at all times" without having to pass through a body of representatives or even through a cumbersome voting machinery. But, of course, the "mechanical difficulties...of working such a method of government are obvious." Bryce wrote:

> How is the will of the majority to be ascertained except by counting votes? How, without the greatest inconvenience, can votes be frequently taken on all the chief questions that arise? No country has yet surmounted these inconveniences.

Today, some 90 years since Bryce wrote *The American Commonwealth*, not only can public opinion be frequently ascertained but the technology also exists to determine it instantaneously. Whether or not this latter capability has strengthened democratic systems is a matter of controversy.

Survey research is a science that provides a variety of tools for discovering, manipulating, and organizing data in many fields, both academic and commercial. Most often, the general public comes into contact with just one aspect of survey research—the public opinion poll.

Public opinion polling is not an easy term to define, although the method is used in all social sciences today and in most related professions. Some define and use the term narrowly to refer to a method of data collection, while others think of opinion polling as an analytical tool. It has been used for evaluation of the past and for prediction of the future. It can provide accurate information about the present; however, when it deals with the past, the data are sifted through and censored by human memory which often is not too reliable.

It has been estimated that $4 billion are spent each year in the United States on all kinds of survey research, an industry that employs millions of people directly or indirectly and has had an impact on every American. The capital investment for survey research in staff, training, equipment, supervision, and interviewing is large. An army of trained interviewers is constantly on call to undertake projects.

The public which reads newspapers and magazines, watches television, or listens to the radio is very familiar with the political opinion poll. Somewhat less frequently, the public also sees or hears new analyses of various social, political, or economic issues that are based on "climate of opinion" surveys.

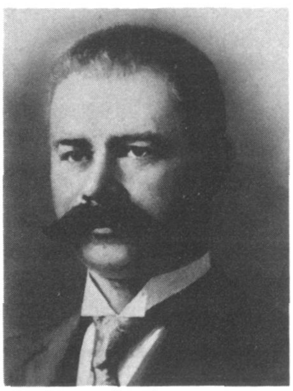
It would be misleading, however, to believe that survey research is used primarily to provide the press with an assessment of the standing of political figures with the electorate, or even to determine how the public feels about the issues of the day. Survey research is also an important tool of the United States Census Bureau, which uses it to collect information on population characteristics, employment statistics, commercial indices, and other social indicators. Market and advertising research would be impossible without sample surveys. Radio and television ratings; sociological and anthropological studies that, among other things, influence the distribution of welfare funds; educational progress reports; economic analyses; the climate for military recruiting efforts; and public health information—all use the data gathered in survey research. In short, whenever it is necessary to gather data about people in the aggregate and their relationships to one group or another, survey research is one of the major scientific tools that makes this possible.

Survey research has many antecedents. As a means of gathering data about the demographic characteristics of people in a social setting (that is, their sex, education, income, or age distributions), surveys go back through history. Surveys were done in the past because society needed the facts to take appropriate actions, and they are done today for the same reason.

Careful descriptive surveys were conducted in the late eighteenth and early nineteenth centuries in Europe and especially in England. John Howard gathered a systematic and objective body of facts in the 1770s that led to prison reform in England. Henry Lytton Bulwer and Frederic LePlay described French social and economic conditions in the early nineteenth century, and Henry Mayhew gave detailed descriptions of various segments of London society in the mid-1880s. The studies of these men were not based on scientific samplings nor were they a complete census, but they did point to the value of social surveys, even when the samples on which they were based were not truly representative of the population.

The application of representative sampling to social surveys followed the industrial revolution and urbanization. With larger numbers of people living in one place, it was particularly important, when laws and regulations were proposed or made, to know as rapidly as possible who would be affected and how.

Before 1880, when the population of the United States was less than 40 million, census enumeration of the entire population took 10 to 20 months. After 1880, with a population of over 50 million, more census-takers were added, and the count was completed in 1 month. Still, the analysis of the census of 1880 took 7 years. Herman Hollerith, an employee of the Census Bureau, helped to halve this time by suggesting that a card with holes punched in it that had been used for some time by weavers to produce patterns in cloth be adapted to the tabulation process. (Hollerith's work is also mentioned in the chapter, "Computers and Semiconductors.")



Herman Hollerith
(1860-1929)
IBM Archives

But even the use of the card was too slow for many needs. Sampling of representative parts of the whole was adopted by several individuals and governments as a compromise between accuracy and speed or, alternatively, economy. Sampling made it possible to get information faster and cheaper than through a census of the entire population. In England and France, for example, sampling of sorts was used in the mid-eighteenth century even before regular censuses were introduced. By the end of the nineteenth century, the pressures for fast and cheap population data had reached the point where sampling for demographic information had become a necessity.

Opinion sampling was a different matter altogether. Only a long-shot gambler would try to predict the outcome of an election on the basis of a straw or "trial" vote, but this is just what the editors of the *Harrisburg Pennsylvanian* and the *Raleigh Star* did. They asked their readers to indicate how they would vote in the elections of 1824 and based their predictions on this unscientific approach. Their chances of

One of the early punched-card tabulating machines.
Smithsonian

Herman Hollerith's machines, used in the 1890 census, were featured on the front cover Scientific American August 30, 1890.

Smithsonian

a correct prediction were not very good. Although the *Raleigh Star* correctly anticipated that Andrew Jackson was in the lead, he received nowhere near the 80 percent vote he received in the straw poll. Nevertheless, by the end of the nineteenth century, such straw polls had become common in both the daily newspapers and magazines.

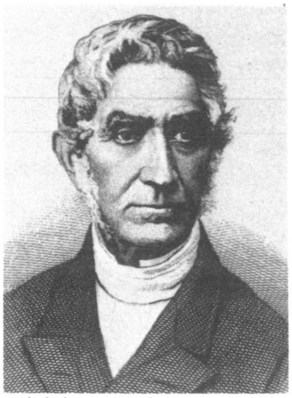
With a very few exceptions, scientific opinion polling did not start until the 1930s when it was introduced in both the marketing and the broadcasting industry. Sampling has its basis in the mathematical theory of probability. Once sampling's scientific validity was understood, its social, political, and commercial value became obvious. From then on, social scientists were encouraged to work on improving the sample reliability and to expand its area of usefulness by developing or adapting statistical tools that permitted more sophisticated analysis.

World War II gave another strong impetus to the development of survey research. The Surveys Division of the Office of War Information conducted surveys of overseas civilian morale, while the Research Branch of the Division of Information and Education in the War Department helped to train, indoctrinate, and evaluate the military forces.

Since then, the development and availability of high speed computers to collate the data, more sophisticated mathematical techniques to analyze what the data mean, and vastly more advanced collection techniques have allowed the science of sample surveys to operate with a much firmer scientific base and with increasing reliability. The applications of survey research have been varied and broad. Most people meet with it more frequently in connection with public opinion surveys. But the same principles of sampling apply to the many other uses of survey research—whether it is in the collection of data for economics, legislation, education, anthropology, or any other field that involves examining the distribution of characteristics in an aggregate.

MEASUREMENT

Asking questions is fundamental to social surveys. Statisticians of the early nineteenth century used questions as an instrument with which to measure social phenomena. But they also found that, despite careful definition and classification of the event (or object) to be measured, subsequent observations of the event (or object) were not necessarily an exact replica but only a reasonable facsimile. The prototype of this facsimile, called the "average man," was described in 1848 by Adolphe Quetelet, a Belgian mathematician, astronomer, and social



Adolphe Quetelet
(1796-1874)
Library of Congress

scientist. He wrote that individual observations will show deviations from the average, but the greater the deviation the less frequently such observations will occur, distributing themselves around the average in the form of a "normal" (bell-shaped) distribution.

Quetelet believed that the concept of normal distribution could be applied to all aspects of the study of society. Mathematics, he decided, was the only way to turn the study of society into a science, because mathematics can focus on what is common to objects. If one could generalize about a branch of knowledge to the point where one could classify things under common headings, it would be possible to reach an advanced level of that science.

But to be able to classify things under common headings, uniform methods of data collection, tabulation, and presentation of findings are needed. He applied this precept in this measurement of the physical characteristics of people (for example, the relationships between their heights and weights), and he insisted that one could do the same for moral and intellectual characteristics.

The English scientist Francis Galton, best known for his pioneering work on heredity, carried Quetelet's theories one step further. He was concerned about the way people were often classified as either ugly or beautiful, honest or dishonest. In his study of eugenics, he found that Quetelet was right; whatever a person's characteristics were, they were normally distributed around a mean of the characteristic. In other words, people had a "central tendency" of beauty or honesty. The further one deviated from that tendency, the fewer people there were who could be said to be that extreme in the characteristic. To discover this, Galton had to devise fairly unambiguous measuring instruments or questionnaires and a rating scale.

The first mailing of a survey questionnaire was initiated in the 1780s by a wealthy Scottish landowner, writer, and parliamentarian by the name of John Sinclair. He sent out 881 questionnaires to the Scottish clergy. The questionnaire was divided into parts dealing with geography and natural history of the parish; age, sex, and occupations of the population; and farming and minerals. He also introduced another survey research technique, the follow-up letter, still used with mail questionnaires. It has been found that follow-up letters are crucial to improving the representativeness of a mail sample.

One other branch in the lineage of survey research is the concept of attitude. In 1848 Quetelet tried to relate various population characteristics, such as age, to the probability that an individual might have criminal tendencies. He called it a criminal penchant or propensity. He said there was a distinction between an "apparent propensity" and a "real propensity." The former is what appeared on the surface; it could be observed and calculated. The "real propensity" is the underlying reason for the "apparent propensity"; it cannot be observed, but only inferred from a person's behavior. Real propensities were what were later referred to as attitudes.

One of the first attempts to measure attitudes through a large-scale survey was by Adolf Levenstein, a self-educated German laborer. Between 1907 and 1911, he sent out 8,000 questionnaires to miners, steel workers, and textile workers, asking them about their opinions on issues such as their political and economic hopes and aspirations, their religious views, satisfaction with their work, their drinking habits, and their cultural pursuits. German sociologist Max Weber prevailed upon Levenstein to publish his findings and suggested ways for him to tabulate and analyze them. Weber was stimulated to think about the application of survey research to the building of theory in the social sciences.

The Discipline of the Scientific Method

Much survey research is done to determine the extent of a particular type of behavior (such as reading a particular magazine), attitude (such as favoring a bill to permit smoking of marijuana), or characteristic (such as country of origin) for a given population. Most newspaper polls and political opinion surveys are of these types. The findings lead nowhere beyond providing an answer to a question. If the question is irrelevant, the answer is irrelevant. Furthermore, if the question is a fixed alternative type of question, offering the respondent a choice of only a few possible answers, the results of the survey are limited by the imagination and ability of the survey designer and such other things as the nature of the problem. As an example, imagine that researchers were living in a culture in which it is believed that illness is caused by a person's being possessed by devils. If they were to design a study to determine the cause of a particular type of illness, they would try to analyze the kinds of devils that were in people with different types of illness. Because they already "knew" the general cause of illness, the researchers would have no need to examine any other possible causes; people do not ask questions that do not occur to them.

The concept of the controlled study is basic to experimentation. (A controlled study in the social sciences is an experiment in which one-half of the subjects are exposed to some treatment, for example, a message intended to persuade them in some way, while a parallel (control) group is not exposed. The purpose is to see what effect the expo-

Blaise Pascal (1623-1662)
Library of Congress

sure has had by comparing the attitudes of the two groups after exposure.) Physical scientists such as Blaise Pascal had been using controls without calling them that since the Renaissance. (Pascal's contribution to the development of computers is included in the chapter, "Computers and Semiconductors.") John Stuart Mill developed the logic of scientific experiments in 1843. In 1900 Karl Pearson contributed the idea of a hypothesis, which is a suggested relationship between two variables that needs to be tested. But modern concepts of experimental design in social and natural sciences are attributed to the British statistician and geneticist, Ronald A. Fisher, who developed them between 1919 and 1930.

Hypotheses are both a necessity and an albatross around the neck of progress in the social sciences. Without them, researchers have no way of deciding what questions to ask in their surveys. A properly stated hypothesis indicates what questions are relevant. On the other hand, once a hypothesis has been stated, it is the only trial relationship that is tested. If it is wrong, the experiment must be done again, knowing only that the first experiment did not confirm a particular relationship. An example might be the relationship between the reading of magazines and education. The hypothesis might be that the more education people have, the more likely it is that they will read magazines. But the real reason that people tend to read magazines may be that magazines give them information they can use in their jobs. In other words, it may be occupation rather than education that is the crucial variable. It is possible that education has little to do with magazine reading but, by coincidence, happens to be tied to occupation. Occupation is known as an intervening variable in this case because, without its intervention in the right direction, the level of education would not predict magazine reading correctly.

Structured Versus Unstructured Interviewing

Interviewing for surveys in the nineteenth century was largely unstructured. Questions could be explained by interviewers in their own words and could be expanded. Questions were not always phrased the same way in face-to-face interviews. To collect demographic data, highly structured questionnaires were not crucial. But in searching for something that is not physically observable, such as a person's attitudes, the only thing that could be held constant was the question. If the questions were not the same, the answers were not really measurable. Unstructured questioning has it uses, however.

Unstructured surveys are exploratory. An example: the United States exhibit in the 1967 World's Fair in Montreal had not been a success. So for the fair in Osaka, Japan, a means to make the exhibit more appealing had to be found. It was decided to approach the problem through an unstructured survey. A set of questions was prepared, and trained psychologists were put to work interviewing Japanese who had been in the United States for no longer than 6 months. They were asked what were the most interesting things they had experienced when they first arrived in the United States, about what things they had written home, and what specifics they wished friends in Japan could see. The unstructured interviews produced a set of ideas that were then worked into a structured (standardized) questionnaire to be used in a survey in Japan. The structured survey established the degree of interest in particular types of exhibits, while the unstructured survey provided the more creative ideas for inclusion in the structured survey. All forms of unstructured interviewing, however, require later standardization by the researcher for the purposes of measurement—that is, the researcher must state the ideas that are common to the various respondents, although each respondent may have used different ways of stating them.

Structured interviews permit measurement of reactions to an unvarying question. If an average response to a given questions is conceived, then it should be possible to measure the amount that a given population varied from that average response. The stimulus is the question; the response is the answer to the question.

Plumbing the Human Mind

One problem with survey research is that it must deal with "self-report." Survey researchers have long been aware of the fact that human memory is imperfect and that people either consciously or unconsciously tend to distort their responses. Even factual questions, such as, "Do you read comic strips regularly, sometimes, or not at all?" or, "How often do you read editorials in your newspaper?" are not answered accurately, sometimes quite unconsciously. The tendency on the part of most people is to provide the socially desirable answer. Some of the social desirability is eliminated in mail or self-administered surveys. One technique that has been developed to counteract the tendency is to introduce the question with: "Most people, as you know, occasionally take things that do not belong to them...." or "Some people believe...while others believe...." Respondents can thus align themselves with one group or another and not feel like misfits.

Another problem of self-report is known as response set. Some people have a consistent tendency to agree with everything. Others tend to be very negative and disagree with everything or to respond to every question on a scale by picking the extreme choices. A survey research technique that has been developed to reduce the effect of agreement or disagreement is to write the questions that might pose problems in a positive vein for half of the sample and negatively for the other half. Thus, half the sample is asked: "Would you say you agree or disagree with the statement: 'Most people are dishonest?'" The other half of the sample would be asked the same question, but "honest" would be substituted for "dishonest."

For the most part, survey research depends on verbal responses. Exceptions are such involuntary measures as galvanic skin responses on which the use of lie detectors is based or measuring the contractions and dilations of the pupil of the eye, a technique which is used in some market studies. Fortunately, people usually tend to tell the truth because of moral persuasion, because they find it easier than telling a lie, or because they are not involved enough to want to lie. It is when people are motivated to disguise the truth that the survey researcher must be most careful. No satisfactory method has yet been devised for assessing the probability that a respondent is distorting answers, although there are some techniques to reduce nonresponse or distorted responses in the case of sensitive or difficult to answer questions.

Projective Techniques and Role Playing

Most people are reluctant to answer sensitive questions about themselves. To overcome this reluctance, survey researchers use projective techniques which ask respondents how they think "others" or "most people" would answer a given question. By projecting the respondents' own behavior to "others," they avoid embarrassment but reveal their own attitudes and behavior.

During the first decade of the twentieth century, the Swiss psychiatrist Carl Jung began developing one form of projective testing—word association studies. Another major contributor to projective tests was the Swiss psychiatrist, Hermann Rorschach. His tests are often referred to as the "ink-blot" test because subjects are asked to react to a spot on a piece of paper which resembles an ink blot. The Rorschach test was the basis of yet another form of projective testing occasionally used in survey research. This other kind of test, the Thematic Apperception Test, was first developed by Harvard psychologist Henry A. Murray in 1935. It involves showing respondents a picture and asking them to use their imagination and, through free association, to answer questions about it. As an example, in a village in Southeast Asia an interviewer shows respondents a picture of a family sitting around a radio set. "To what program are they listening?" the interviewer asks. The respondents know that most people listen to forbidden Communist broadcasts, but the villagers would be foolish to admit to a stranger that they do so. Yet this is a make-believe picture. There is no problem. So a respondent says the family is listening to a Communist broadcast.

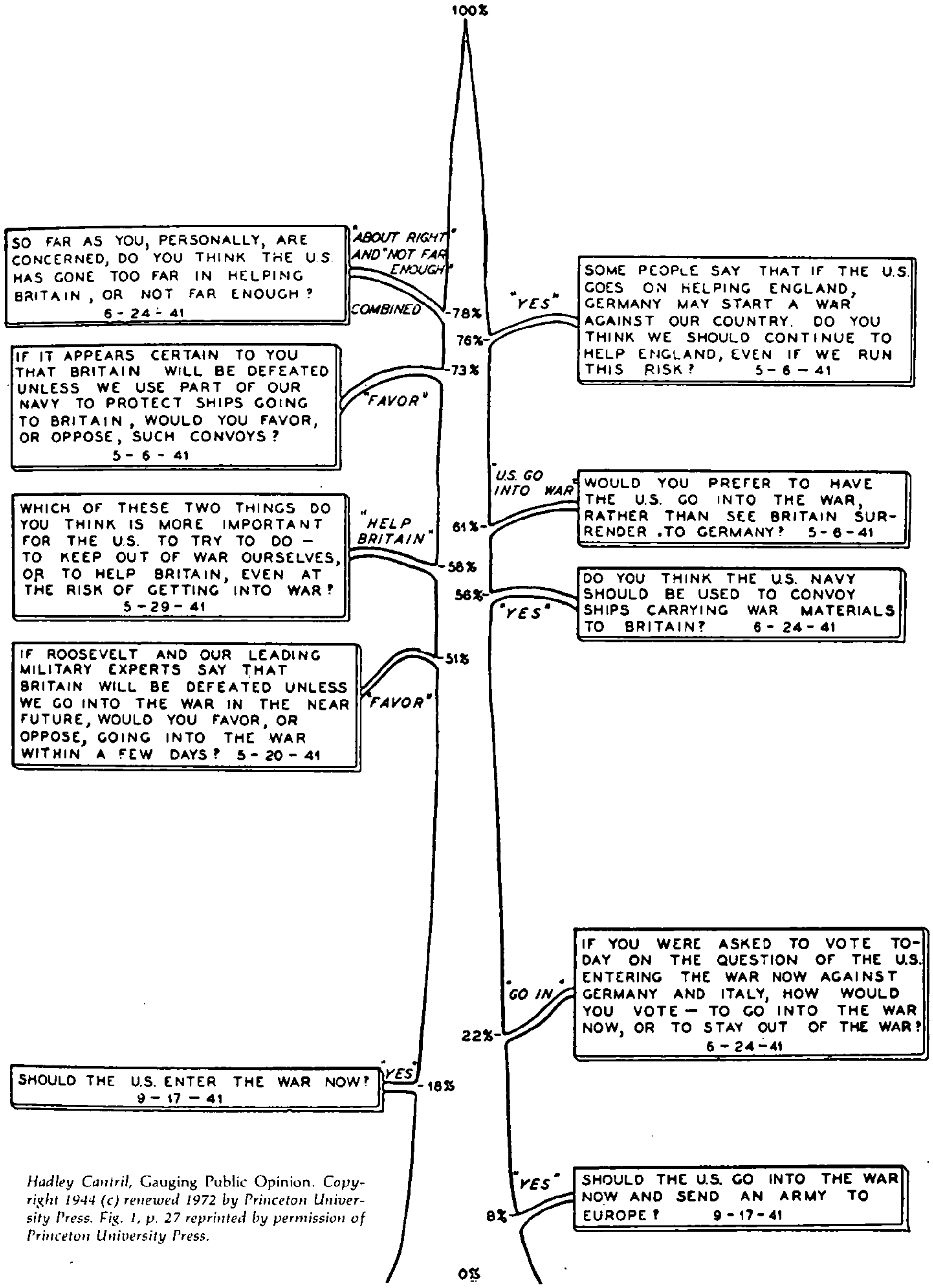
Unfortunately, most respondents approach a survey interview as though it were an intelligence test. Philip E. Converse, a survey research specialist at the Survey Research Center of the University of Michigan, found that the average respondent, when faced with an opinion question dealing with an unfamiliar topic, is reluctant to say, "I have no opinion—I have never thought about those things before you came and never will again after you leave." Instead, an opinion will be ventured.

In a 1956 study Converse tried to relieve the respondents who might not have an opinion of any compulsion to answer by introducing the question with: "Of course, different things are important to different people, so we don't expect everyone to have an opinion about all of these. If you don't have an opinion, just tell me that." In spite of this, follow-up studies and analysis showed that only an additional 27 percent gave no opinion upon this special invitation. It was assumed, based on previous experience, that about 9 percent would have opted for a "don't know" response anyway. That left 64 percent who chose to offer an opinion. Of these, 45 percent could be shown by comparing and analyzing their responses in follow-up interviews to have no opinion.

The 1956 study by Converse and other research have shown that only around one in five members of the public has an informed opinion on matters dealing with public affairs. Unless an opinion is grounded in strongly based attitudes, it is not a reliable guide to future bahavior.

Question Wording

The problem of how to word a question to obtain the desired information is a difficult one, and there is no fool-proof rule to guide the survey researcher. As Humpty Dumpty put it in *Through the Looking Glass and What Alice Found There:* "When I use a word...it means just what I choose it to mean, neither more nor less." This, of course, is an exaggeration because culturally and socially shared meanings of words exist. But the fact that one can maneuver people into responding positively to a most serious question affecting everyone in the population was illustrated by American social psychologist Hadley Cantril. Using the responses to surveys conducted by the American Institute of Public



Hadley Cantril, Gauging Public Opinion. Copyright 1944 (c) renewed 1972 by Princeton University Press. Fig. 1, p. 27 reprinted by permission of Princeton University Press.

How Opinion Varies with Contingencies Involved. During May through September, 1941, surveys showed that between 8 percent and 78 percent of Americans favored involvement in World War II. What made the difference was whether they were asked bluntly, "Shall we send an army to Europe?" or an indirect question such as, "Do you think the U.S. has gone too far in helping Britain or not far enough?"

Opinion and the Office of Public Opinion Research, just before the United States entered World War II, Cantril showed that from 8 percent to 78 percent of Americans favored involvement in the war in Europe.

Note in the figure on page 93 that differences in responses hinge on whether the question is asked directly, by implication, or by a combination of direct and implied statements.

Over the years, certain rules for question construction have evolved. One rule that has been recognized for a long time is that for most purposes questions should be asked in the most simple and economic form possible. The words used should be understood by the person with the least education in the sample. Sometimes the question is a concealed two-part question, as in the following query from a Congressman:

> Health care costs have increased substantially in recent years, to the extent that adequate health care has become difficult for many people. Should the Federal government consider a National Health Insurance program to cover all Americans?

The respondent who is for a national health insurance program, but not for one that covers all Americans, is in a quandary. The question should have been simplified by being split into two separate questions. Split question techniques have the disadvantage, however, of lengthening the questionnaire.

Leading questions are another favorite of those who are more interested in the public relations effect of their survey than in the climate of opinion. This example is from another questionnaire:

> Are you in favor of a tax cut, even though it might lead to a greater Federal deficit?

Some questions are simply ambiguous. They mean different things to different people, and the survey researcher cannot tell which was meant, as in the following question:

> Do you believe that gun control can solve the crime problem?

The meaning of "gun control" may range for some people from simple registration to confiscation of all firearms; for others "solve" may mean either reduce or eradicate. Adding all of the responses to this question is like adding apples and oranges.

The split-ballot method was developed in 1940 by the American Institute of Public Opinion to test the effect of wording differences. The method, which is often used by market researchers, as well, involves using two versions of a question with different halves of the sample of respondents. The following is an example that shows the emotional content of the italicized words:

A. Would you favor *adding a law to** the Constitution to prevent any President of the U.S. from serving a third term?

B. Would you favor *changing** the Constitution to prevent any President of the U.S. from serving a third term?

Responses	Percent Response for "Adding"	Percent Response for "Changing"	Percent Difference
Yes	36	26	10
No	50	65	15
No Opinion	14	9	5

*Not emphasized in original survey.

The order in which questions are asked also can have an important effect on the response, as in the following:

A. Do you believe that anybody should be allowed to speak on any subject she or he wants to, or do you think there are times when free speech should be prohibited?

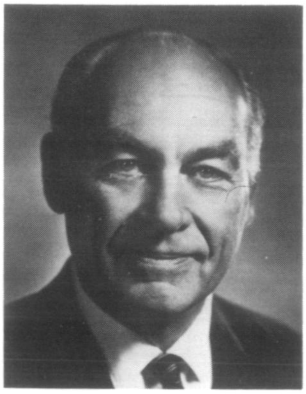
B. Do you believe in free speech to the extent of allowing Communists and extreme rightists to hold meetings and to express their views in the community?

Because the researcher has little control over the order in which questions are read in the case of self-administered questionnaires (as in a mail survey), questions that may be affected by their sequence are difficult to ask, and must be masked or camouflaged.

In the modern questionnaire there are essentially two kinds of questions: open-ended and closed-ended. In open-ended questions, respondents fill in their responses (or an interviewer does it for them) in their own words. This is useful when the researcher has no idea what the answers might be. It is used extensively in a pilot or preliminary study when information is being collected about scope or issues. Open-ended questions have their drawbacks, however. If the interview is being conducted by an interviewer, it takes a fast writer or a tape recorder to take down everything the respondent says. If the interviewer summarizes, the danger of injecting personal interpretations arises. If the questionnaire is self-administered, some respondents may tend to skip open-ended questions because they are too much bother. If everything is taken down verbatim, there is also the problem of coding later. Someone must decide how to classify the answers under a limited number of categories, otherwise answers cannot be reduced to numbers for purposes of comparison.

Closed-ended questions, also called fixed-alternative, multiple-choice, or cafeteria questions, precode the responses. This means that the researcher lists the choices for the respondent to consider in the question. For example:

> Compared with 5 years ago, would you say your attitude toward your local hospital has become better, has remained about the same, or has become worse?



George Gallup
(1901-   )
The Gallup Organization, Inc

Closed-ended questions have the advantage of focusing all respondents' attention on the same alternatives. The responses can thus be more easily tabulated. On the other hand, the alternatives presented force the respondent into an artificial mold. The true opinions of the population may lie elsewhere.

George H. Gallup, in 1947, developed a procedure that he called the quintamensional design to overcome some of the disadvantages of open-ended and closed-ended questions. This method begins with an open-ended general knowledge question to determine whether respondents are aware of the question. Then it is followed by another open-ended question in which respondents are asked for an opinion about the issues in their own words. Next comes a question with alternatives on some specific part of the issue, and respondents choose among them. This is followed by a question asking "Why?" to permit respondents to elaborate on their choice. Finally, there is a closed-ended intensity question to determine how strongly the views are held.

Because of the possibility that trends in opinion may be studied at some future time, questions are often worded in exactly the same way as they were in some past survey. Any changes in wording would make it difficult to determine whether it was the change in wording or a change in attitudes that influenced the results of a survey.

When responses to questions are tabulated, it should be remembered that they are responses to words that have only approximately the same meaning for the various respondents. This is one of the reasons why sample surveys do not necessarily reflect the "true opinions" of the population.

Another reason is that surveys reflect a situation at a given time; yet they are often projected into the future. When asked how they will vote tomorrow, people may oblige and tell the interviewer how they think they will vote tomorrow. But often implicit in their response is the warning, "I hope you won't hold me to it." Survey research may be an important guide to the future, but that does not necessarily make it a good guide. Yet it is a principle, both in everyday life and in scientific endeavors, that the best evidence available should be the guide. Survey research has provided increasingly better evidence on which to base decisions on certain topics.

SAMPLING

Its Origins

In 1912 a group of city officials with a problem in Reading, England approached Arthur Bowley, a part-time lecturer in economics and statistics at the local University Extension College. They were concerned about the condition of the working class in the city. But Reading could not raise the money for a social survey of the magnitude of earlier, more sweeping studies. Bowley was just the person to tackle this problem. Although his education had been in mathematics at Cambridge, his interests were in economics and social issues. He had heard of the great debates at the International Statistical Institute conventions at Bern in 1895 and Budapest in 1901 on the "representative method." He had also heard that Anders N. Kaier, chief of the Norwegian Bureau of Statistics, had used "representative sampling" in gathering labor statistics in 1891. So he undertook to use the limited funds available to prepare a report on "Livelihood and Poverty" in Reading. It was a landmark study for sampling techniques.

In Reading, it was economy that forced the use of sampling. In 1923 in Japan it was the need for speed that precipated its use. An earthquake in Tokyo destroyed part of the tabulation of the Japanese census. To get the results out by the following year—by age, sex, size of households, and other demographic characteristics—Japanese officials took a sample of 11,000 by picking every thousandth household of the 11 million in the census. When the tabulations of the entire census were later completed and compared with the sample results, it was found that the sample had provided very accurate data.

George H. Gallup, who started a commercial polling organization in the early 1930s, showed that as long as a sample was representative, it did not have to be very large to reflect consistent results. Thus a survey of public opinion on prohibition provided the following results:

	Percent Favoring Prohibition	Percent Opposing Prohibition	Percent with No Opinion
• First sample of 442 respondents:	31	62	7
• Adding second sample of 442, thus making 884 respondents:	29	63	8
• Third sample of 443, making 1,327 respondents:	30	63	7
• Fourth sample of 1,258 making 2,585 respondents:	31	61	8
• Fifth sample of 2,670, making 5,255 respondents:	33	59	8
• Sixth sample of 2,998, making 8,253 respondents:	32	60	8
• Seventh sample of 4,241, making 12,494 respondents:	32	61	7

True, it is not known what the distribution of opinion regarding prohibition was in the population as a whole. But there is no reason to believe that further sampling—to the point where the total population is approached—would have caused percentages to fluctuate any more.

It took a long time for the laws of probability and sampling, which had been used in connection with natural and physical observations, to be applied to social phenomena. Jakob Bernoulli, a seventeenth-century Swiss mathematician, physicist, and theologian, demonstrated in 1713 that the colors of a sample of marbles drawn from an urn would be approximately proportionate to the colors of all the marbles in the urn if the marbles in the urn were randomly distributed. This illustrated empirically, or through experiment, what Dutch mathematician Christian Huygens and French mathematicians Pierre de Fermat and Blaise Pascal had figured out theoretically 50 years earlier. (Huygens' contribution to seismic exploration is mentioned in the chapter on that subject.) Bernoulli stressed that a substantial number of observations would have to be made for the sample to be accurate.

Bernoulli suggested that sampling might be applied to economics. About 100 years later, in the early 1800s, a French mathematician and astronomer, the Marquis de Laplace, tried to figure out the probability that jurors might err in arriving at a verdict. His compatriot mathematician, Simeon Denis Poisson, worked on the same problem and, in 1837, developed the bases for the application of statistics to the social sciences. But, it was not until the end of the nineteenth century that British scientist Francis Galton began to apply statistics to biological problems.

It had been the common wisdom that "like begets like." Galton set about proving this axiom empirically and statistically. To do this, he measured and weighed fathers and sons. He found a high correlation between the heights and weights of fathers and sons, but he also discovered that fathers who were very much taller than the average tended to have shorter sons, while shorter than average fathers tended to have taller sons. He called this "regression" to the mean (or average).

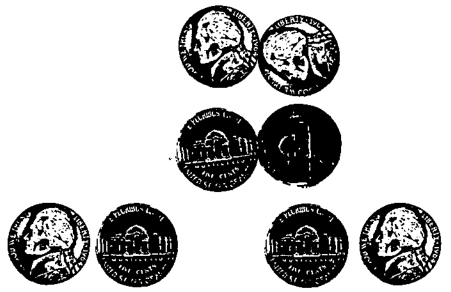
Bowley, the man from Reading, pointed out in 1912 that what could be done with biological problems could also be done with sociological ones. He stressed that, while the number of variables encountered when dealing with human beings may be unlimited, what is important is the variability of single attributes of these human beings, such as how their opinions vary on a given topic or the extent to which their opinions differ from the average. Actually it was this measurement and observation of variance that first led to the study of probability in physical observations.

Around the middle of the seventeenth century, a Frenchman by the name of Chevalier de Méré called on Blaise Pascal to help him improve the odds on his gambling. Fascinated with the challenge, Pascal worked on the problem and checked it out with his friend, Pierre de Fermat. One of the first things Pascal did was to toss coins. He then worked out how many chances he had of getting a given number of heads and tails in any number of trials.

Being a mathematician, Fermat immediately saw that the number of times each chance is likely to occur could be figured out mathematically, thus:

Tossing a single coin one can get a Head or a Tail

With two coins, one can get



With three coins, one can get

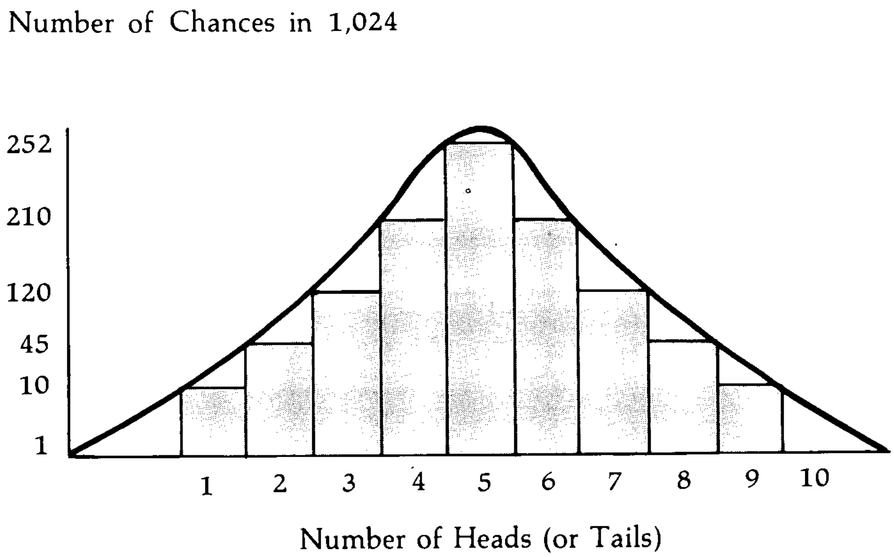
Pascal prepared a triangle of chances based on up to 10 coins that looked something like this:

Number of Coins		Sum of Chances
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256
9	1 9 36 84 126 126 84 36 9 1	512
10	1 10 45 120 210 252 210 120 45 10 1	1,024
	10 9 8 7 6 5 4 3 2 1 0	

He read the last line of his triangle as follows:

> If 10 coins are tossed once or 1 coin 10 times, there is 1 chance in 1,024 of getting all heads or tails; 210 chances in 1,024 of getting either 6 heads and 4 tails, or 4 heads and 6 tails; etc.

Graphing this, he found the following kind of figure that shows the decreasing probability of getting mostly heads or mostly tails:



Abraham DeMoivre, a French mathematician refugee in England, in the early 1700s smoothed the resulting curve by increasing the number of cases (or coins) to a very large number, and in the latter part of the century the French astronomer-mathematician Marquis de Laplace introduced further improvements. In the early nineteenth century, the German mathematician and astronomer, Karl Friedrich Gauss, used the Laplacian curve to develop the concept of standard deviation and of standard error from the mean. To provide any data, however, that are susceptible to mathematical manipulation, researchers have to agree on just what it is they are measuring.

Observation of nature had shown that no two things, at least among biological and social phenomena, were exactly alike. In other words, generalizations here are theoretically impossible. But if minor differences could be ignored, it was argued, categories or classes of things and ideas that are approximately alike could be lumped together.

Since, however, nothing is an exact replica of anything else, if something is named, say, "leaf" and thought of as a ideal, every other leaf after that deviates from the ideal. Statisticians refer to the deviation as "error." The second leaf is not identical, but it is still close enough to be called a leaf. What Gauss did was to define and calculate the average error that is likely to be found in anything in nature. He called the average error a standard deviation, which is a form of average deviation from the mean (or average) of all observed cases.

An example of how mathematicians calculate this variance or deviation from the ideal might help to clarify the theory behind it. If the term "weekly income" is used at the example, a sample of this income might be the weekly paycheck over a period of 5 weeks, thus:

First Week	$120.00
Second Week	80.00
Third Week	70.00
Fourth Week	100.00
Fifth Week	80.00

The total income for the 5 weeks is $450, and the average, or mean, income is $450 divided by 5, or $90. Note that the mean is an ideal or representative weekly income. None of the weeks in this particular sample shows that amount. Each week is in "error" to some extent—that is, if $90 were the ideal representative of the term "weekly income," Gauss figured out how to define how much each was in error. He also noted that the mean, by definition, is a midpoint. If the deviations from the mean, the pluses and minuses, were to be added, they would total zero. To eliminate the zero, he squared the deviations, since this would not affect the *relative* distance of the deviations but would take care of all the minus signs, thus:

	Observation	Mean	Deviation from Mean	Squared Deviation from Mean
First Week	120	90	+30	+900
Second Week	80	90	-10	+100
Third Week	70	90	-20	+400
Fourth Week	100	90	+10	+100
Fifth Week	80	90	-10	+100
Total	450	450	0	1,600

The average squared deviation from the mean is 1,600 divided by 5, or 320, and this Gauss called a variance. If the variance were "unsquared" or the square root taken, the result would be about 18. He called the number calculated in this way the standard deviation.

When Gauss examined the Laplacian curve (which has since been renamed the Gaussian curve), he deduced its mathematical form and noted that the curve had the shape of a bell. He figured out that about 68 percent of the area under that curve was within one standard deviation from the mean; 95 percent of the area was within two standard deviations; and better than 99 percent of the area was within three standard deviations from the mean. To illustrate his discovery, the measurement of the results of intelligence quotient (IQ) tests can be used. For this example, average IQ can be given a rating of 100. Obviously, not everyone has an average IQ. Some are above average and some are below. How much people's IQ deviates from the average can be calculated by subtracting 100 from each person's IQ rating and working out the standard deviation. If it is discovered that the standard deviation is 15, according to Gauss' approach about 7 out of 10 people would have IQs that measure no less than 85 and no more than 115—that is, one standard deviation (15) on each side of the mean (100).

Furthermore, 95 percent of people would be no more than two standard deviations (15 multiplied by 2, or 30) removed from the mean, putting them between 70 and 130 in IQ, and 99 percent of people would be no more than three standard deviations from the mean (45), placing them between 55 and 145 on the IQ scale.

A great deal of work has been done to determine the optimal number of responses needed to assure accurate responses to survey questions. It has been found that even if a survey is repeated a number of times using different samples of the same size, the proportions for a particular question will vary from sample to sample—but not by much. These changes in response rates from sample to sample are known as the sampling error. To halve the error, a sample four times that size is needed. This is where a compromise must be made between accuracy and economy. Two questions must be asked. The first is, how much error can be tolerated? In a survey dealing with attitudes toward a political candidate, a fairly large amount of sampling error may be acceptable. On the other hand, if a survey were made to determine whether a drug will have harmful side effects, a much smaller error would be tolerable.

The second question is, how important is it to be certain that this sampling error will not be exceeded? One standard deviation covers close to 70 percent of the curve. Actually, some 32 percent of the sample would lie outside this area. Thus, if the sampling error tolerated is 3 percentage points, and a survey showed 60 percent favoring a candidate, the true population percentage might be anywhere between 57 and 63 percent. It would be about 68 percent certain that this error of 3 percentage points is not exceeded. Two standard deviations would provide over 95 percent confidence or only a 5 percent chance of the error being greater, and three times the standard deviation yields better than 99 percent assurance.

Populations and Sampling Frames

Imagine a carload of grain. To determine how much impurity is in the grain, a random handful could be taken to count the number of foreign particles in that sample, assuming that the entire sample is well mixed. If the handful contained 1,000 individual seeds and 10 of the particles counted were foreign particles, that would represent a ratio of 10 to 1,000, or 1 percent. There is no need to examine the entire carload. This one handful would give a fairly good estimate of the incidence of impurities.

The real world, however, is not that homogeneous; populations are not uniform. In the 1870s the German statistician and economist

Wilhelm Lexis developed a theory for cluster sampling. That method involved drawing samples from population clusters which tended to have more homogeneity internally, or within a cluster, than there was between the clusters. He devised methods for drawing subsamples from each of the clusters so that the final sample was representative of the entire population.

A major problem with sampling is defining a population so that there is no doubt whether a given unit or element is or is not a member of that population. Unless the population can be defined precisely, samples may not be representative.

Having defined a population—for example, all citizens who are registered to vote and are not traveling abroad or in jail or in a hospital—it is necessary to find a sampling frame from which to draw the sample. This sampling frame, a list or map of all the units in the population, must be comprehensive and up-to-date. A telephone directory, often used as a sampling frame, illustrates many of the problems of defining the population. If the population is defined as "telephone owners," a directory does not include unlisted numbers, nor does it include all names of individuals who share a single telephone. Moreover, it includes many names of people who have moved or whose numbers have been changed as well as business telephone numbers that are not identifiable with a particular person or persons. If they are so identifiable, that person often is listed again at his or her residence giving the person more than one chance of being selected. In short, defining the population unambiguously and finding a sampling frame to fit the definition are difficult and important challenges to sampling and survey research. Failure to meet these challenges has caused some massive errors in the past.

For 20 years, the *Literary Digest* had been conducting presidential election surveys based on mailing lists it had started collecting in 1895. In 1936, 10 million ballots based on telephone and car registration listings were mailed out and 2,367,523 straw votes were returned. These telephone subscribers and automobile owners gave 54 percent of their vote to Alf Landon and 41 percent of their vote to Franklin D. Roosevelt. The electorate, on the other hand, gave Roosevelt over 62 percent of their votes. Obviously, the *Literary Digest* voters, in spite of their numbers, were not representative of the electorate. The magazine's straw votes were discontinued that year—as was the *Literary Digest* itself.

Two approaches to achieving representativeness have been attempted: quota sampling and probability sampling. Quota sampling involves dividing the population into relevant groups based on factors such as age, educational level, sex, geographical location, and race. Quotas are set for each of these groups in proportion to their prevalence in the population. Put simply, if 20 percent of the population has a college education and 40 percent a high school education, while another 40 percent has less than a high school education, then those are the proportions by which the sample is drawn.

Quota sampling has a number of shortcomings. Accurate and up-to-date figures on population demographics are essential to setting the correct quotas. But, more importantly, the relevant characteristics for a particular survey must be known—something that is normally better known after the survey than before. Furthermore, if the interviewers are given free rein as to whom they include in the sample to make up the quota, there is always the chance of picking people who are most like the interviewers—that is, more congenial. Quota sampling, therefore, is being used less and less in serious surveys, especially in view of the fact that the error cannot be estimated when it is used. The probability that characteristics will fall within a calculated margin of error is based on the assumption of randomness in the choice of subjects, and pure randomness cannot be assumed in quota sampling.

There are two approaches to probability sampling. The purest and best is based on a listing of the entire population, from which a random sample is drawn by using a table of computer-generated random numbers which indicate the element or person to be picked. Numbered listings of the population for use as a sampling frame are hard to obtain or do not exist. So a second method, based on the geographical location of the elements of the population, is used most of the time.

Area sampling, as it is called, might use a map over which a grid is placed, marking out cells that are picked randomly. Within each cell or cluster thus chosen, everyone can be interviewed, or the clusters can be enlarged and subclusters can be picked for inclusion in the sample.

The problems of selecting respondents; of dealing with refusals, "not-at-homes," wrong addresses, and wrong identifications; and of making substitutions plague survey researchers to the present day. One basic rule of sampling is that for a sample to be representative of the population from which it is drawn, it must be truly random, which means that every element in the population must have an equal or a known chance of being selected, and every combination of elements must have an equal or a known chance of being selected—an always difficult and sometimes impossible condition to ensure.

While sampling theory has changed little in the past 30 or 40 years, the methods of securing and analyzing the data and of reaching the respondent have changed significantly. Even more significant is the discovery of what will and what will not work. In the 1920s and early 1930s, mail surveys were popular. But by the mid-1930s, commercial and academic pollsters had decided that the personal interview was superior, although far more expensive. The Voice of America has used radio announcements where access to the population is difficult, if not impossible, especially in sparsely populated African countries. These call on listeners to write in, giving some information about themselves, in exchange for which they will participate in a drawing for a prize.

One obvious advantage of mail surveys is relatively low cost. Another is that the interviewer does not affect the responses by influencing the answers. The disadvantages are numerous. The respondents are self-selected, that is, the respondents decide whether they will or will not return the questionnaire. The bias is toward people in the upper educational and socioeconomic levels, who are more likely than those in the lower levels to answer their mail. The sequence in which the questions are answered cannot be controlled, and that may influence the responses. There is no way of telling whether the questionnaire was filled out by the addressee or by some other person. And, most important, it is not possible to tell how the views of those who did not reply differ from those who did.

For many years it was the consensus that the only truly scientific surveys were those conducted through personal interviews. Their advantages remain numerous. They are still the best way to exercise full control over the sample, although reaching the selected individuals may be difficult at times and may take a number of callbacks. They are certainly the best method for long interviews.

Because more than 94 percent of households in the United States have telephone service, telephone surveys have gained in acceptance since the 1930s and 1940s. Today, it is possible to reach almost all segments of the population and to reach them quickly and relatively cheaply. In most cases, the interviews must be kept short—perhaps no longer than 20 minutes—but recent studies suggest that once the respondent is on the telephone, the length of the interview ceases to be a major problem. The biggest problem is with complex and long questions, which are hard to administer over the telephone. Questions that involve the ranking of alternatives, or even selection among a large number of choices, simply cannot be administered.

Panel studies involve recurrent interviews with the same sample of individuals. They are frequently used for consumer studies and for broadcast ratings. But they have also been used to study how the electorate makes up its mind, as in the seminal Erie County, Ohio, surveys by Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet in 1940. Panels are most useful in studies of trends and causes of change, but they suffer from a high drop-out rate and from the dangers of contamination. Contamination, which is extremely difficult to determine, occurs when responses to later interviews are influenced by conditioning caused by earlier question; those who are exposed to the questions are more likely than others to notice relevant items in the mass media or to discuss the topic with friends.

Broadcast ratings are often done through panels. The A. C. Nielsen Company has used a meter that is attached to television or radio sets and checked periodically. Nielsen and the American Research Bureau also use a diary method, which involves leaving diaries with a panel of subjects who fill them out as they listen to a radio or watch television. The problem with meters is that they merely record whether the set is on and to what channel it is tuned, not who, if anyone, is listening or viewing. And it is impossible to ensure that diaries are kept when memories are still fresh.

Coincidental telephone interviews are done by C. E. Hooper, Inc., Trendex, The American Research Bureau, and others. They involve calling at the time the program is being broadcast to determine how many people are listening and who these people are. One problem is that the person most interested in the program is least likely to answer the phone. "Day recall" telephone interviews conducted by the Sindlinger Company, among others, overcome this problem. People are asked which programs they heard the previous day. The weakness is the shortness of many people's memory. Personal interviews covering the previous 24 hours, conducted by Pulse, Inc., have the same shortcomings.

In recent years, electronic devices have been introduced both to assist in the interviewing process and to obtain instantaneous feedback from viewers of a television program. One such polling device received wide publicity in July of 1979 when it was tested for instant feedback from 6,000 to 8,000 viewers of a presidential speech to the nation. The polling system works in conjunction with the Warner Communications Company's QUBE cable television service in Columbus, Ohio. (QUBE is also discussed in the chapter, "Synthetic Fibers.") Its 30,000 subscribers have control boxes on their television sets with numbered

Viewers of QUBE cable television in Columbus, Ohio are able to communicate their responses directly to the studio, thereby providing instant feedback to questions asked by the station.

QUBE

buttons which register at the studio. Percentages reflecting subscriber responses are calculated instantaneously, thus making it possible for the audience to provide and to receive immediate feedback. Because subscribers are representative of nobody but themselves and tend toward the upper socioeconomic levels and because the respondents who choose to participate in any particular survey are not even representative of the subscribers, results of such polls are no more than interesting at this point.

Ultimately, the objective of a sample survey is to make meaningful statements about the population from which the sample is drawn. Theoretically, one can say that proportions found in a random sample will be the same as those to be found in the population as a whole, within calculable margins of error.

Political Polling

In 1932 George Gallup conducted a survey for his mother-in-law, who was running for Secretary of State in Iowa. She won the election and thus became the first woman Secretary of State in Iowa. The first paid-for survey commissioned by a politician was undertaken in 1946, when Jacob K. Javits had the Roper organization undertake a survey of the 21st Congressional District of New York. But, by the mid-1970s, three out of four congressional candidates were using professional pollsters, at least to write their questionnaires or to design their samples, if not to conduct the entire survey.

The information gathered most frequently in the first poll in a political survey deals with issues important to the electorate. At least as many questions are asked to determine the relative standing of candidates. Somewhat fewer surveys examine the candidate's image and the strengths and weaknesses of the opponent, and a relatively smaller number of surveys also check on the effectiveness of the various media with the electorate. The costs have already run into the hundreds of thousands of dollars and continue to rise.

Since 1960 when pollster Louis Harris conducted the first private poll for a presidential candidate (John F. Kennedy), no candidate for national office has run without the help of surveys. One estimate indicated that by the mid-1960s, 85 percent of the winning senatorial candidates and more than half of the elected members of Congress had used polls.

The accuracy of reputable polls has increased tremendously since the fiasco of 1948, when almost all pollsters predicted that Thomas E. Dewey would defeat Harry S. Truman in the presidential election. The average combined error of final congressional and presidential election polls of the Gallup organization from 1950 through 1978 has been about one-half of 1 percent.

But not all pollsters have proved to be as accurate as the Gallup polls, and some are unscrupulous in their use of survey research. Charles W. Roll, Jr. and Albert H. Cantril wrote in *Polls: Their Use and Misuse* that when the *New York Daily News* showed Charles Goodell running a poor third in the Senate race, his staff was contacted by a polling firm which asked if Goodell would be interested in purchasing a poll that showed him to be ahead. Another firm, bidding for a senatorial candidate's business, offered to provide two surveys: one for fund-raising and publicity purposes and another to show the true standing of the candidate.

Dan Nimmo reports that candidates themselves often misuse survey findings. In November of 1967 Lyndon B. Johnson's staff circulated a private poll showing the President leading all contenders in a "bellwether" area. Archibald M. Crossley, who had conducted the survey, revealed that his findings had been distorted—that the survey had, in fact, been done in a traditionally Democratic county in New Hampshire.

Government Surveys

Governments have always had a need for information about the people they governed, and, in one way or another, they have always managed to collect this information. The concentration of large numbers of people in urban areas during the industrial revolution increased this need and called for more detail. Rising immigration led to the collection of city-by-city data on country of origin with the census of 1850; beginning in 1870, educational reforms required detailed data on occupation and education. The need for accurate population data kept increasing, and more and more studies were done. But these were generally hasty and poorly conceived reports that ignored the availability of sampling techniques being developed at the time.

It was not until the census in 1940 that the United States Census Bureau added a 5 percent sample survey of households to include questions on employment and income. Since then, the Census Bureau has been adding topics until today it has quarterly rotating panels to estimate population characteristics such as social and occupational mobility, the labor force, income levels, ownership of major appliances, housing plans, and consumer buying habits.

During World War II, the Research Branch, Information and Education Division of the U.S. Army, did about 300 studies on some one-half million soldiers. These studies helped in policy decisions affecting troop morale, training, racial desegregation, and demobilization. They also provided a generation of students and scholars with a gold mine of data and methodological models for studies in social and communication research.

After World War II, a number of other government agencies added survey operations, including the Bureau of Labor Statistics of the Department of Labor, the Congressional Research Service of the Library of Congress, and the International Communication Agency, which conducts opinion and media surveys in many foreign countries. These surveys are used in foreign policy decision-making in evaluating the impact of its own media and cultural activities.

Government-sponsored and government-conducted surveys provide data for many commercial and social uses. The surveys are essential to the proper functioning of a modern society. But the data they generate must be accurate—a goal which can be achieved only if they are based on scientific survey practice. Inaccurate data that masquerade as accurate and scientific are not only misleading and useless but are also potentially dangerous.

Marketing and Media Surveys

Among the early names associated with business and marketing research was that of Archibald M. Crossley. Starting as research manager of J. H. Cross advertising agency in Philadelphia after graduating from Princeton, he later became assistant director of research for the ill-fated *Literary Digest* polls. In 1926 he set up his own business as a market researcher, using surveys as a business aid for newspapers, advertising agencies, and corporations. He was one of the first to sample public reactions to brand names, and when radio broadcasting became widespread, he pioneered in surveys of radio advertising impact and listening habits. One of the most widespread uses of survey research today is for rating radio and television programs. Commercial ratings of network radio programs were first undertaken by the C. E. Hooper Company in 1935.

Then, the A. C. Nielsen Company, founded in the early 1940s, invented its mechanical recording device, the audimeter, that was attached to radio receivers to monitor the station and the amount of time a set was tuned in to it. Nielsen bought the Hooper Company in 1950. By the mid-1970s Nielsen was offering an "instantaneous Nielsen" service by connecting some 1,200 sets to a central computer for overnight tabulation of audiences.

Pulse, Inc., entered the broadcast rating business in the mid-1940s using the personal interview method. In 1952, the American Research Bureau added the diary method (and later telephone coincidentals and recall surveys). Today the American Research Bureau is called Arbitron. Other companies, such as Trendex and Sindlinger and at least 50 more, have been doing both national and local rating surveys since the 1950s. The accuracy and the methods of the broadcast ratings generate a great deal of controversy. In 1961 they were the subject of hearings by Representative Oren Harris' subcommittee of the House Interstate and Foreign Commerce Committee, which reported many shortcomings—some of which have now been corrected.

Today, much of the estimated $4 billion that is spent annually on survey research goes into consumer and media audience research. Corporate market surveys help estimate what types of products consumers are likely to buy. Time series analyses investigate past buying behavior to try to predict future patterns of buying. Market testing is done with the help of sample surveys to determine whether a new product has a chance of capturing a large enough segment of the market to make marketing it worthwhile. Market share analyses help a manufacturer determine how successful a product is in competition with similar products of other manufacturers. Packaging research uses surveys to study the most appealing ways to package a product.

ANALYSIS

Probably the most progress in survey research has been in the sophistication of the analysis of data. Computers provided a major impetus to this development. The earliest survey researchers were satisfied with mere frequency distribution, percentages, and a simple breakdown of the data by demographic characteristics. In short, they reported who gave what answers, what proportion gave each answer, and what the proportions were in each town, age group, sex, income level, and so forth. While the analysis of the data is far more sophisticated than it was 30 years ago, validation of the findings still needs improvement. Not much attention was paid to the validity of the findings until the 1930s, when American psychologist Samuel A. Stouffer began to ask whether a relationship existed between what was measured and the intention of the investigator. The validation of survey data is difficult. The data are valid only if they measure specifically what they were intended to measure. Some kinds of surveys are tested for validity on the basis of whether they correctly predict a future event—an election or some other test of behavior, such as what proportion of the public plans to buy a certain product. Yet it cannot be determined whether the survey proportions were correct and remained constant until the election or purchase; whether they were incorrect, but the proportions changed and happened to coincide with the election or purchase when those occurred; or whether publication of the results turned them into a self-fulfilling prophecy.

Studies by survey specialists Hugh J. Parry and Helen M. Crossley have shown the validity of survey findings to be somewhat disappointing. They showed that almost 25 percent of a sample of the electorate claimed they had voted in the 1944 elections when they had not. Other studies show that more people tend to "remember" having voted for the winning candidate than actually did so.

One method developed to validate survey data is to compare responses with such known data as age or sex. Theoretically, if the sample is truly random, all characteristics of the population—including age, sex, education, and whatever opinions may be held by the population—would be reflected in the sample in the same proportions as they exist in the population as a whole. Thus, if the verifiable characteristics are correct, everything else is considered correct—a leap of faith. Another validating method involves correlating the index or measure used with some other accepted measure. An index of religiosity might be the number of times a person goes to church or it might be a minister's judgment. If the number of times someone goes to church happens to coincide with the minister's judgment of her or his religiosity, then the index has "criterion validity." A third validation method of survey research is known as construct validity. Conservatism, for example, is a construct that may be measured by a number of indices, including social and economic level and political persuasion. Survey findings regarding the conservatism of a population have construct validity if indices of conservatism are included and measured by the survey.

Validity is related to reliability, but it is not the same. A watch may be highly reliable, meaning that it is consistent and does not lose or gain any time. But it may not be valid: that is, it tells the wrong time, for example, Paris time rather than New York time.

There is no doubt that survey research can define, describe, and capsulize the prevailing attitudes, behavior, and characteristics of a population. Without such techniques it would not be possible to make group comparisons or predictions or to develop norms (standards against which to evaluate individuals or groups). But it must be remembered that estimates are what are involved—albeit the best estimates available.

It has been estimated that some 2,000 survey research organizations conduct surveys, and there are many times that number of businesses that conduct their own studies. The technology is deceptively easy to learn, and the rewards are high. The United States Congress often requires that at least a part of any money that is appropriated by the executive branch for projects of all types be set aside for evaluation—frequently survey research is used for this purpose.

The American Statistical Association has been concerned about this requirement and has sponsored a study to assess survey practices. Of 26 federally sponsored surveys, 15 did not meet their objectives because of poor design or sampling methods; technical flaws, such as high refusal rates, not-at-homes, etc.; or poor supervision of interviewers or of coding and punching of data. Seven of the 10 nonfederal surveys also failed to meet ASA objectives.

The social science community is concerned about a variety of problems affecting surveys. One of them is the growing number of unqualified firms and individuals who are conducting them. Another is the fact that the mass media occasionally either publish survey results without giving adequate information about the method of data collection or draw inappropriate conclusions from the results. A third is that unscrupulous salespeople occasionally insinuate themselves into a home on the pretext of conducting a survey and then switch to a sales pitch. Serious deficiencies remain in the scientific quality of some survey information, in the practices of the industry, in an over-reliance on surveys, and in the demands it places on a heretofore cooperative public.

A review of the positive and negative contributions of survey research to American society suggests that it is, and has been, important as a planning resource, a managerial aid, a citizen information base, and a means of constructing and testing social theories. Like most other sophisticated tools in use today, survey research must be continuously refined if it is to keep pace with the society it was designed to serve.

Pesticides and Pest Control

Since the first cave dweller took a swing at a flying gnat, pests have been a problem. Some are merely nuisances; others pose serious threats to crops, forest products, livestock, and human health.

For thousands of years efforts to manage agricultural productivity were, at best, only moderately effective and resulted in relatively low and unpredictable production levels. Before the twentieth century most successes in achieving higher agricultural yields had been piecemeal. Farmers failed to consider or were unable to defeat the entire range of crop-limiting factors at once. If a farmer used an improved crop variety, for instance, the crop was often inadequately fertilized or watered or was destroyed by pests.

For years the only controls that kept plant and animal pests from completely overwhelming the world's crops were natural ones—genetic resistance to pests, weather adverse to pests, natural predators, and parasites. Plant and animal pests could, and sometimes did, destroy entire harvests. The Bible recounts plagues of locusts that "covered the face of the... earth" and ate "every herb of the land and all the fruit of the trees." In the 1840s in Ireland the late-blight of potatoes (*Phytophthora infestans*) brought about the deaths of more than 750,000 people and compelled the emigration of at least a million more. More recently, grasshopper hordes decimated vegetation in the prairies of North America, leaving behind a land stripped bare. These "plagues" were, fortunately, the exception; but when they occurred, they were devastating.

Because helicopters can spray smaller and irregularly-shaped areas more easily than a small aircraft can, they may be used to apply pesticides. This helicopter is treating an orange grove near Lake Wales, Florida with a fungicide.
USDA

It was not until about 40 years ago that scientists working in chemistry, plant genetics, entomology, engineering, and a host of other fields began to develop truly effective methods for improving and ensuring crop yields. One of their most dramatic successes was in giving the farmer new weapons—chemical pesticides—to use in the age-old annual battle against those organisms that threatened their crops and livestock.

The term "pest" means different things to different people. To many, it means only harmful insects and rodents; but the term also includes noxious organisms, such as mites, nematodes, weeds, and certain vertebrates. In addition, essentially all organisms that cause plant disease, whether fungi, bacteria, nematodes, or viruses, are considered pests.

Still, there is a continuing semantic problem about what is and what is not a pest. Butterflies are beautiful, but their larvae may be serious plant pests. The wheeling flight of a flock of starlings may be a beautiful sight to a poet or a city dweller, but to a farmer with a newly-seeded field it is a nightmare. Birds are also considered pests when they destroy crops such as cherries and strawberries. Clearly, honeybees are normally highly desirable insects, but they may be pests to individuals who are particularly sensitive to their sting or to others who find themselves unwittingly surrounded by a swarm of bees.

Circumstances of quantity, location, timing, and culture also affect the status of an organism as a pest. Bermuda grass may be useful in erosion control and as a source of hay, but it can be a damaging weed in agriculture or a problem in home gardens. Dogs are prized as pets, but stray dogs may be legitimately classified as pests.

Most pest control measures, however, are directed at species that are generally recognized as pests. The control achieved can be considered a benefit to society, whether it results in reduction in the transmission of malaria or plague or in the number of grasshoppers; higher yields of sweet corn, wheat, or apples; or less damage to roses in a garden.

While the exact amount of world-wide crop loss attributable to pests is unknown, it is estimated to be about one-third of the potential food production—food for about 1 billion people. Rice has the greatest loss from pests—possibly as high as 40 to 50 percent of potential production.

The magnitude of the pest problem can be appreciated even more when it is recognized that the amount of land now under cultivation in the United States is half again as much as would be required if there were no pest-induced losses. The crop-related energy needs and amounts of labor, capital, fertilizer, and other requirements also are correspondingly greater. The total costs are huge.

The United States, particularly, has seen a shift from small farms that grew a wide range of crops to large, specialized farms that engage in monoculture—the intensive production of single crops. This development has provided pests with a highly concentrated food source, in and around which they can breed and feed with unprecedented ease. There is little doubt that monoculture has led to a greater need for pest control. Without monoculture, however, the population would have been more permanently bound to the land, and technological development probably would not have occurred as rapidly as it has.

Civilization must defend itself not only against organisms that destroy buildings, clothing, crops, and stored food but also against certain pests that adversely affect health. If epidemic typhus, malaria, trypanosomiasis, encephalitis, and other diseases that have caused widespread death and debilitation are to be avoided, defense against disease-transmitting creatures must be carried out continuously. Pests that bother people and higher animals by buzzing, crawling, stinging, biting, or sucking blood must also be controlled.

This photograph, taken in 1946, shows DDT being applied to kill corn borers in an El Paso, Illinois cornfield. The photograph also illustrates the agricultural practice of monoculture.
USDA

Use of the pesticide DDT (dichlorodiphenyltrichloroethane) brought about a dramatic change in the world. It was less than two generations ago that victims of typhus and malaria were dying by the thousands in the United States and by the millions around the world. By helping to control the spread of insect-borne diseases, DDT saved millions of lives and prevented millions of people from becoming sick during the early years of its use.

Without the development of DDT and its continued use against the pests which carry typhus and malaria (body lice, rat fleas, ticks, and *Anopheles* mosquitoes) it is likely that the ravages of these two killer diseases would still be a major factor in mortality for the United States or for other countries where they have been essentially eliminated. DDT and other pesticides also have been used to control the spread of yellow fever, bubonic plague, certain types of encephalitis, and a host of other diseases.

In the early 1960s, Americans became much more aware of the uses and the dangers of pesticides. They grew especially conscious that some of these toxic chemicals persist in the environment and accumulate in certain plants and animals and even in human beings. Rachel Carson's book, *Silent Spring*, spearheaded their interest. A report prepared by President Kennedy's Science Advisory Committee in 1963 also contributed to this awareness and called for a reduction of the use of chemical pesticides. These publications helped set the stage for rapid expansion of public involvement and concern. Since that time both the scientific community and the public have become increasingly aware of the need to weigh the risks of this technology against its benefits. A substantial effort has been made to limit the use of persistent pesticides and to replace them with ecologically safer chemicals or other methods of pest control. Alternative methods are also being sought because, over time, pests frequently develop tolerance, resistance, or immunity to particular chemicals or control measures. Cultural methods (crop rotation, tillage, proper timing of planting, and fertilizer and water management) are now being used more extensively. Other approaches to pest control include biological control (regulating pests through their natural enemies, including vertebrates, insects, mites, fungi, bacteria, and viruses), the use of behavior modifying substances (pheromones), sterilization techniques, eradication programs, and the use of synthetic hormones to disrupt critical stages in life cycles.

Current and potential problems with chemical pesticides demonstrate the need for increased research to develop more and better alternative techniques. Until those techniques are developed, however, some pest problems exist for which chemical controls are the only available answer.

Because of the adaptability of pests, possible hazards to the environment, the economic impact of pest combinations, and concerns about the use of certain chemical pesticides, there is great interest in developing and refining an approach to pest control which would promise both more stable crop production and protection and the least possible risk to human beings and to the environment. That approach is called Integrated Pest Management (IPM).

The term, "Integrated Pest Management," has been defined in a variety of ways. It is described in a 1979 report of the Office of Technology Assessment (OTA) of the Congress of the United States as "an all-inclusive concept that should be applicable to all pests (weeds, plant pathogens, nematodes, vertebrates, insects, etc.). However, terminology, control tactics, and strategies vary among disciplinary groups, so that it is difficult to arrive at a definition completely appropriate to all interests." In that report, OTA uses the definition that IPM "is the optimization of pest control in an economically and ecologically sound manner, accomplished by the coordinated use of multiple tactics (Tactics are the specific methods used to achieve pest control. These include chemical pesticides, pest-resistant varieties, cultural practices, biological control, and others.) to assure stable crop production and to maintain pest damage below the economic injury level, while minimizing hazards to human beings, animals, plants and the environment."

A key to understanding the role of pesticides in society today, both in regard to the benefits of increased crop yields and improved public health, is to recognize that pesticides have not eliminated the problems of pest control. They have allowed society partially to control the damaging effects. If pest management were to stop, the life-threatening situations and crop and livestock destruction caused by pests would rapidly reappear.

A HISTORICAL VIEW OF PESTICIDE DEVELOPMENT

The practice of applying various substances to plants, soil, livestock, and sometimes even to people for the control of pests goes back many years. Fine dust or ashes sprinkled on leaves of crops to combat leaf-eating worms were among the earliest pesticides. Salt was occasionally used to discourage the growth of weeds among certain salt-tolerant crops. A story about how South American natives killed fish by using the root of a certain plant (*Derris elliptica*) led to the discovery of rotenone, an insecticidal compound. This substance was applied as a dust long before the science of chemistry permitted identification of the active complex chemical ingredients of this natural product. The insecticidal properties of a powder derived from dried flowers of the daisy-like *Chrysanthemum coccineum* plant were discovered in the early

1800s when it was observed that insects avoided this plant. Since then, the specific chemicals, pyrethrins, which gave the flowers their insect-killing power, have been extracted and have been synthesized.

Farmers used arsenicals such as Paris green as early as 1865 to combat insects which were damaging potatoes in the Rocky Mountain region. Fumigation with hydrogen cyanide and lime sulfur washes were used in the 1880s to fight insect pests in California. Although these inorganic compounds were better than nothing, they had to be used in high doses and even then were sometimes not very effective. They were also highly toxic to human beings.

Legend has it that the development of one chemical pesticide was brought about by a group of thirsty French thieves. To stop their nocturnal raids into the vineyards of Bordeaux to steal grapes, vintners developed a combination of copper sulfate and lime. The grapes were sprayed with this "Bordeaux Mixture" to keep the "human pests" away. The mixture left an unappealing blue residue, copper hydroxide, and made the grapes toxic if they were not thoroughly washed. The human pesticide worked. One day an alert grower noticed that the mixture also discouraged growth of mildew, so after the thieves were no longer a problem the vintners continued their spraying—and a chemical pesticide was born.

In 1919 E. B. Blakeslee found that paradichlorobenzene could be used to protect clothes against moths; later it was found that this same chemical was also effective against peachtree borers (*Synathedon exitiosa*).

Horse-drawn power dusters were used in the early 1900s. In the operation depicted, sulphur or arsenate of lead is being dusted on trees as the wagon is slowly pulled down the road.
USDA SEA

At about the same time, it was discovered that some chemical compounds had valuable insecticidal as well as fungicidal properties, and systematic searches for such laboratory-produced pesticides were undertaken in earnest.

The early basic research on pest-control techniques was done mainly by biologists who concentrated on classifying particular types of pests and studying their life cycles, engineers concerned with the control of mosquitoes by draining swamps, and chemists who worked to improve the few chemicals known to be pesticidal. In the early part of the twentieth century, scientists were searching specifically for parasites and predators of important pests. Biological control was almost all they could offer because there were few effective chemical pesticides. Eventually chemists became involved on a more practical level. They improved formulations, refined the products, and identified chemical structures. Through research they explored methods of synthesizing chemical pesticides and of improving ways to apply them. Closely related to this latter research was the start of research on the behavior of emulsions, spray droplets, dusts, and other physical states important to the application of pesticides. Studies such as these involved physics and engineering as well as chemistry and biology.

Basic studies of kinetics of charged particles and emulsion structure in the 1930s allowed scientists to prepare chemicals that would permit formation of an emulsion suitably stable for long-term storage. Emulsions were needed in which chemicals remained uniformly suspended when diluted in the spray tank before application, but in which the droplets of the emulsion would break immediately upon contact with the sprayed surface so the active pesticide could adhere. Formulations and equipment for applying pesticides were modified to increase the concentration of the active ingredient in the spray tank, thus lowering the volume of diluent which had to be transported and applied.

The basic patent for the pressurized gas aerosol method of applying substances was granted to two chemists in the United States Department of Agriculture—Lyle Goodhue and W. Sullivan. Aerosols later became popular for many household uses.

In 1874 a German chemist, Othmar Zeidler, synthesized DDT and filed it away as a curiosity. In 1939 that curiosity provided pesticides' quantum jump when another chemist, Paul Müller, at J. R. Geigy, S. A., in Switzerland, discovered the use of DDT as a pesticide (for which he received the Nobel Prize in Medicine and Physiology in 1948). Besides being credited with saving millions of lives which might have been lost to insect-borne typhus and malaria during and after World War II, the discovery of DDT set off a new effort of basic research into the chemical control of pests.

This new effort was led by the establishment of the United States Office of Scientific Research and Development, a wartime group of outstanding scientists from many disciplines. Among the responsibilities with which this group was charged was that of finding new and improved methods of controlling typhus, malaria, and other infectious diseases. This challenge required the cooperation of organic chemists to develop new compounds, physiologists to determine the mode of action for pesticidal chemicals, and toxicologists to determine hazards involved in the use of the new chemicals.

By the end of World War II, this team effort proved so successful it was rapidly expanded to apply to agricultural pests, including weeds. Basic studies on the interaction between chemical structure and biological activity led to discovery of such new forms of pest control as the hormone-like activity of 2,4-dichlorophenoxyacetic acid (2,4-D) on plants, natural and synthetic insect hormones, and pheromones.

For a number of years after World War II, DDT and other chlorinated hydrocarbons were dominant in the insecticide market. Over time it became apparent that such stable substances can be dangerous because they accumulate and migrate in the environment. Moreover, pesticide resistant mutations, which sometimes seemed to cause even more destruction than their ancestors, were beginning to emerge.

Resistant strains result from selective survival of individuals having a genetic constitution which provides immunity. Although such resistance has been described since the early 1900s, it began to increase in the middle of the 1940s when large scale application of pesticides became common practice. In some cases insects have developed resistance to as many as four different toxic chemicals or to other pest control measures.

Alternative pesticides which did not pose threats to the environment were sought. As a result, organophosphorus insecticides (such as malathione and parathione) and carbamate insecticides (such as Sevin® and Furadan®) came into increasing use. The organophosphates were developed from a research base that began in Germany in the early 1930s at I. G. Farbenindustrie. Some of these chemicals are systemic (absorbed by a plant through its leaves, stems, or roots) causing a treated plant to become lethal for a time to insects that eat it. (Nonsystemic insecticides are lethal to insects that touch or ingest them directly.) Organophosphates can be used to control diseasecarrying mosquitoes and animal pests such as fleas, ticks, cattle grubs, and hornflies. The organophosphates remain in the environment for a shorter time than the chlorinated hydrocarbons and, therefore, have limited uses. Despite this relative lack of persistance, some organophosphates have been found to be dangerously toxic to human beings.

Development of the carbamate insecticides began in the 1940s at J. R. Geigy, S. A., in Switzerland. Most of these substances act topically, but a few show a high degree of systemic activity. Although they do not tend to leave harmful deposits on foods, some are harmful to warm-blooded animals. Other carbamates are effective because they act on biological reactions and processes that are common in insects but are not as common in mammals.

Recently, the evaluation of the effects of pesticides has been greatly influenced by the tremendous advances in sensitivity and specificity of chemical analytical technology and techniques. Before the 1950s only relatively high concentrations of pesticides and their metabolites could be identified and measured. Residues of less than one part per million of pesticides in animal or plant tissues were considered analytically negligible because they were regarded as trace residues that could not be measured.

In recent years the development of improved analytical techniques, however, has provided scientists with the ability to analyze and identify certain pesticides and other chemicals in concentrations as low as one part per trillion. (One part in a trillion is the equivalent of about 1 minute in 1,900,000 years.) People have become more concerned about pesticides at these lower concentrations, and some research has shown that, even at low concentrations, significant human risks can result from the persistence of some chemicals in the environment. Although the defenders of pesticides characterize such minute concentrations as "one drop of vermouth in a tank car of gin," others take the position that even these low concentrations of certain substances are unacceptable. Another effect of the drive to achieve pesticide use that is almost or completely risk-free has been to intensify the interest in Integrated Pest Management.

THE DEBATE OVER CHEMICAL PESTICIDES

Just as any technology may yield new benefits—real or imagined—it can also produce new hazards—also real or imagined. A conflict over which hazards and benefits are real and which are imagined seems inevitable.

Pest control by chemicals is a relatively new technology which most people agree should be integrated with biological, cultural, genetic, and other means of pest control. But any evaluation of the use of pesticides must also include an analysis of what happens when no pesticides are used. Obviously, both sides of the pesticide ledger must be totaled before decisions on the greatest net benefit can be made.

Evaluating the trade-offs in pest control methods is often a matter of dealing with a moving target. Circumstances are changing continually. New regulations and technical improvements and discoveries that reduce risks must be part of one side of the equation, while known or potential hazards that are detected in initial safety screening tests become a part of the other side.

Many agriculturists believe that those who are concerned with health must balance the production ledger against the possibility that an occasional person might be injured now or in the future as a result of trace pesticide residues. Those in agribusiness point out that the use of nonchemical methods of pest management alone would result in a severe drop in crop production. Many fear that the public health may be more adversely affected by reduced crop yields and increased incidence of pest-borne diseases than by the occurrence of cancers, birth defects, or other illnesses in a few individuals.

The potential of some pesticides to cause human illness and even death has been recognized for some time by government regulators and the agricultural chemical companies. Directions for use and label precautions in the United States are essentially legal documents; deviation from them is a violation of the law. The time that must elapse following application before harvest or slaughter, the interval required before workers can re-enter a sprayed field, and safe disposal features for used containers are all a part of the precautionary statements. In addition, pesticide products are classified into restricted and nonrestricted groups; only a certified farmer or commercial applicator can use restricted materials. And some pesticides can cause problems even when used as directed.

On the assumption that human beings must protect themselves, their crops, their forests, and their livestock from pests by one means or another, all available information must be evaluated and understood before decisions can be made wisely. Those who began their gardening before the end of World War II will recall the skull-and-crossbones sticker required on packages of arsenical insecticides. Although some people were concerned about residues of these inorganic substances on apples and other fruits, pesticide regulation was minimal. When the

This trap is used to monitor the number of codling moths in an area where sterile males have been released to control the population of this pest.
APHIS—USDA

codling moth (*Laspeyresia pomonella*) developed resistance to lead arsenate, gardeners increased the number of applications to combat the pest with the result that even drastic washings would not remove enough of the pesticide to meet the established tolerance levels. Then, based on a study done by the United States Public Health Service, the Food and Drug Administration raised the tolerance levels for this insecticide. With the introduction of a wide range of pest control materials soon after 1945, however, product registration procedures with requirements for safety and efficacy testing prior to marketing were established. A statement of recommended handling precautions, as well as use directions, was required on the label.

Federal Regulation of Pesticides

During the early 1950s people became more concerned that exposure to certain substances such as cigarette smoke and asbestos might increase the probability of developing cancer. One consequence of this concern was that in 1958 Congress included in the Federal Food, Drug, and Cosmetic Act what is now known as the Delaney Clause. That clause prohibits the use of any food additives that are found to be carcinogenic to human beings or to animals. The ruling applies only to substances that are added purposely to food; it does not apply to natural ingredients of foods, inadvertant or unavoidable contaminants, or to pesticides applied to crops in the field. It does cover pesticides that might contaminate food as the result of pest control techniques during storage or processing of food as well as residues of pesticides if they become more concentrated during storage or processing of the food. The clause neither offers guidelines as to what tests for carcinogenicity are "appropriate" nor requires testing for carcinogenicity, although the requirement is implied.

In 1964 the United States Department of Agriculture decided to refuse to register pesticides for use on food crops unless the Food and Drug Administration established a finite tolerance for a residue of that pesticide on a particular crop or food. This action prohibited the registration of potentially carcinogenic pesticides for such uses. Some experts thought that carcinogenicity testing should be required for all newly registered pesticides. At about the same time, representatives of the agricultural chemical industry urged standardization of testing required for registration of pesticides so they could complete testing without undue delay before seeking registration. In response to their request, in the early 1970s the Environmental Protection Agency (EPA) attempted to develop complete guidelines for testing necessary to register a pesticide. Many tests that could be used for one pesticide, however, could not be used for another; the task was abandoned. EPA then proposed that complete carcinogenicity testing be required for those pesticides expected to result in residues on food, food crops, or feed crops; on those substances suspected of being carcinogenic because of chemical or other similarities to known carcinogens; and for those substances that tested positively in the Ames Test for mutagenicity. (The Ames Test is used to determine whether or not a substance is likely to cause inheritable mutations.)

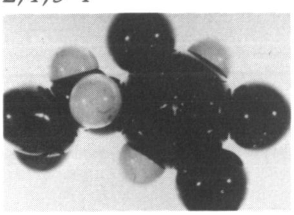
Prior to the Delaney clause, premarketing safety tests for new pesticides did not place major emphasis on testing for carcinogenicity. Subsequently, a wide-range retesting (with laboratory animals) of pesticides already on the market was instituted. Traditionally, in designing long-term feeding tests, toxicologists have used at least one concentration of the substances to be tested that is high enough to induce some effect in laboratory animals. Because many substances being investigated show no effect at doses comparable with those to which human beings are typically exposed, it is sometimes necessary to use relatively high dosages to obtain any effect.

The proper interpretation of data about carcinogenicity from these high-dosage-level tests constitutes one of the most frequently debated aspects of pesticide controversy. One group of people claims that even a single molecule of a carcinogen can, over time, produce cancer if it damages the genetic material of a cell. They theorize that, if tumors in laboratory animals can be produced in tests using grossly exaggerated dosages or using very large numbers of animals at low doses over a long period of time, human cancers can result from infinitely smaller doses. When this belief is added to the ability of chemists to detect toxic substances at concentrations as low as one part in a trillion, to some people the world seems to become a carcinogenic nightmare. Other people are convinced that there are concentrations of most toxic substances below which no harmful effects will occur. Neither group can convince the other.

Where reliable epidemiological data exist, they are useful in determining carcinogenicity. Unfortunately, epidemiology is of limited value in most instances where chemical pesticides are concerned—primarily because, for most chemical carcinogens, it may take as long as 20 years or more between the first chemical exposure and the onset of cancer in human beings. By the time a substance is confirmed as a carcinogen, many people may already have been exposed to it. Epidemiologists also have trouble locating people who may have been exposed years previously to a particular substance. When and if such individuals are found, they often have also been exposed to several other potentially hazardous agents. Thus it is difficult to prove that a pesticide is carcinogenic unless there is a well-defined, easily traceable, relatively homogeneous group of people who have had similar exposure to a particular substance.

In 1972 the Congress passed another piece of legislation which has had a great effect on the use of pesticides in the United States. The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) places the responsibility for proving the safety of a pesticide on those who wish to make them—not on the general public as represented by the EPA.

2,4,5-T



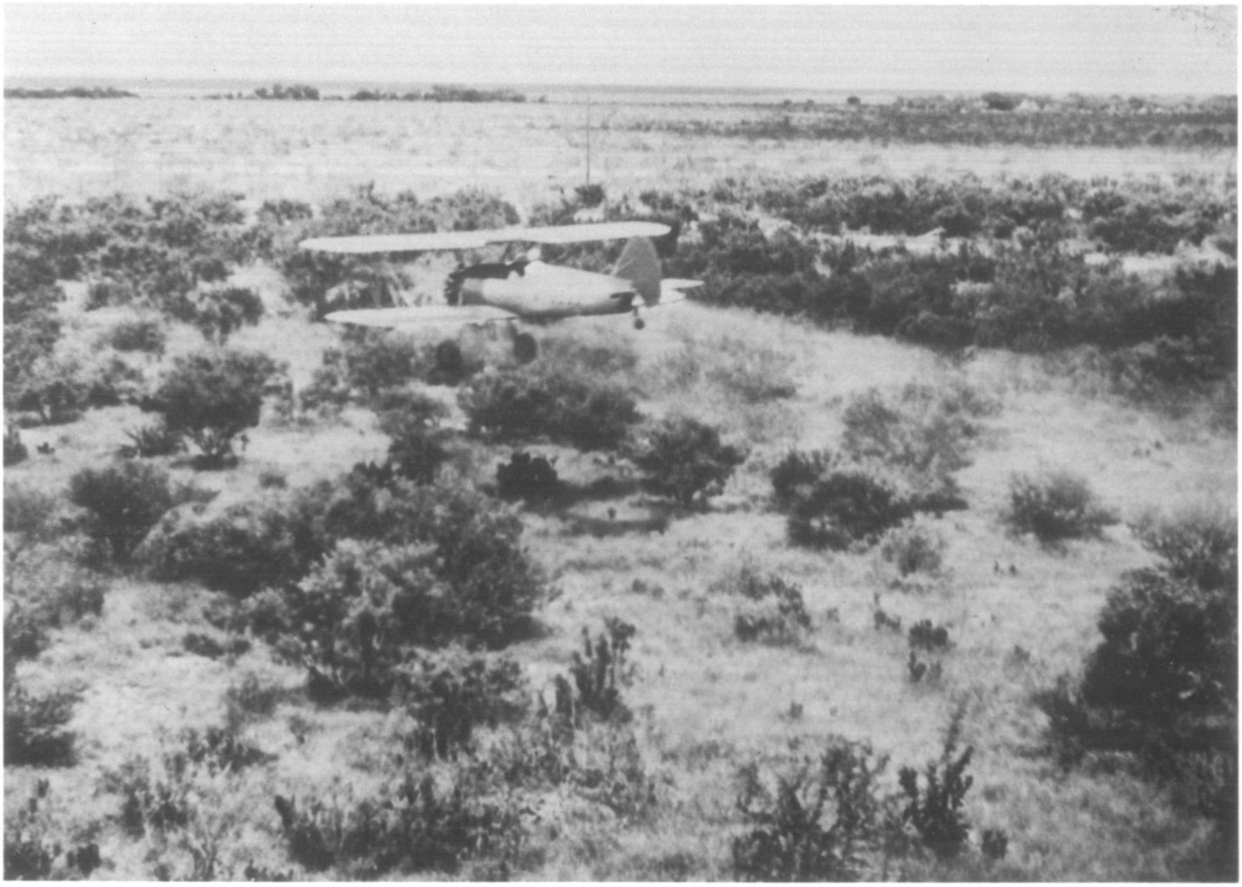
The model represents the chemical structure of 2, 4, 5-trichlorophenoxyacetic acid (2, 4, 5-T). Dow Chemical Corporation

In 1970 the majority of pesticides (an estimated 90 percent) were organic chemicals. Insecticides and fumigants, including nematocides and rodenticides, accounted for about 47 percent of the organics. The other organics were divided between herbicides (37 percent) and fungicides (16 percent). Between 1966 and 1976, the use of pesticides in the United States doubled to more than 600 million pounds. (The use of insecticides and fungicides increased only slightly during this period.) The much greater use of herbicides is responsible for most of this growth. In 1977 herbicides accounted for 58 percent of all pesticide sales in the United States. Although the *total* use of chemical pesticides in the United States is not expected to increase appreciably over the next 15 years, the use of herbicides is expected to accelerate considerably.

The herbicide 2,4,5-T (2,4,5-trichlorophenoxyacetic acid) has been in use for about 30 years in the United States, Australia, New Zealand, Germany, and a number of other countries. Each year about 3.8 million acres in the United States are treated with 2,4,5-T—principally to control brush and weeds on land used for right-of-ways, timber, rice, and grazing. Other methods for controlling these types of vegetation generally are not as effective or economical as 2,4,5-T. For example, if the current use of 2,4,5-T on United States forest land were cancelled and other known silviculture methods were used, it is estimated that the forest industry would sustain a cumulative net income loss of about $800 million over 10 years. If 2,4,5-T could not be used on right-of-ways, it is estimated that expenses for right-of-way maintenance would increase 35 percent (about $35 million) annually.

In this photo, 2, 4, 5-T is being sprayed to control brush and weeds on rangeland. This herbicide has also been sprayed along right-of-ways, on timber, and on rice-growing areas to control the growth of weeds.

Dow Chemical Corporation



The dramatic controversy which surrounds the use of the herbicide 2,4,5-T is one example of some of the complex scientific, political, and economic considerations which can be associated with the use of certain pest-control methods. Late in the 1960s it was reported that high dosages of 2,4,5-T caused birth defects in experimental animals. In fact, it was not the herbicide itself that produced the birth defects in these animals, but rather a highly toxic impurity—the dioxin TCDD (2,3,7,8-tetrachlorodibenzo-*p*-dioxin) (shown on page 129)—that occurs in the formulation of 2,4,5-T. The presence of TCDD is the principal reason for the concerns about the herbicide.

The herbicide 2,4,5-T was a key component of "agent orange," a defoliant widely used in jungles during the war in Vietnam. The discovery that 2,4,5-T had the potential of causing birth defects was instrumental in discontinuing the use of "agent orange" in Vietnam. Currently a number of veterans of that war allege that their exposure to "agent orange" has seriously damaged their health by increasing their risk of cancer and causing them genetic damage. The validity of these charges is yet to be determined. Prior to the late 1960s certain quantities of 2,4,5-T contained up to 30 parts per million (ppm) of TCDD. When the Dow Chemical Company, the principal manufacturer, learned of the high toxicity of TCDD, it changed its manufacturing processes so that the herbicide produced by 1969 contained less than 1 ppm.

In 1970, not long after the ban of "agent orange" for military uses, the United States Department of Agriculture, United States Department of the Interior, and the United States Department of Health, Education, and Welfare jointly agreed to prohibit certain uses of 2,4,5-T. The Environmental Protection Agency (EPA) set a hearing date in 1974 on cancellation of the remaining registered uses of 2,4,5-T.

It had been found that TCDD disintegrates rapidly (that is, within days) in some environments receiving bright sunlight, thereby theoretically reducing the risk of using 2,4,5-T on rice fields, on right-of-ways, and on rangelands or pastures. On the other hand, it was not known how rapidly TCDD breaks down under certain other conditions, such as after being sprayed on forests. Moreover, it was not clear whether or not TCDD would concentrate in animal tissues, as do some other chlorinated hydrocarbons. Furthermore, only a handful of laboratories in the world had the highly trained technicians and sophisticated equipment necessary to detect very low concentrations of the chemical.

By the time scheduled for the public hearings, the most advanced methods of mass spectroscopy had so increased the sensitivity of analysis that as little as 10 parts per trillion (ppt) of TCDD could be detected in some tissues, but the method had not been verified in a sufficient number of laboratories to be useful in the hearing. Without information about human exposure (usually derived primarily through analysis of diet residues, tissues, or milk) there could be no sound evidence of such exposure or, therefore, of risk to human health. In spite of scattered reports of problems in human beings and animals attributed to 2,4,5-T, the public hearings scheduled for 1974 were cancelled because of a lack of sufficient data.

In the meantime, the EPA created a process called the "rebuttable presumption against registration" (RPAR) to help identify pesticide uses which may not satisfy the legal requirements for a pesticide's registration. The procedure also establishes a method for gathering and evaluating information about the risks and benefits of these uses and allows for public participation in the process. In effect, for a pesticide to be registered for a use, its benefits must exceed its risks for that use. The RPAR for all pesticides containing 2,4,5-T products was issued in 1978. During this process, in March of 1979, the EPA announced an emergency suspension of spraying of 2,4,5-T on forests, in pastures, and along right-of-ways. Use of 2,4,5-T on rangeland and rice were not included in this suspension order. The EPA said it did not say "that the health effects in humans are positively proven, or that 2,4,5-T should never be used again." But it did say that "there is sufficient evidence to stop further exposure of the chemical until the issues can be resolved." The evidence (collected during the RPAR review) to which they were referring was, in large part, information derived from the results of animal studies and an epidemiological study.

The principal impetus for the emergency suspension of the use of 2,4,5-T was an epidemiological study said to show a "high probability that the herbicide is linked to human miscarriages in an area where 2,4,5-T is used regularly." The study (called Alsea II because this was the second study conducted in the Alsea Basin region in Oregon) which precipitated the suspension has since been severely criticized by a number of independent experts. Their criticisms include incomplete and inaccurate data on 2,4,5-T usage, inaccurate data on spontaneous abortions and failure to recognize that the monthly variation in rates of hospitalized spontaneous abortions described in this study was no greater than might be expected through random variations. The consensus of those who reviewed the Alsea II study (including a number of scientists at a 1979 conference on 2,4,5-T sponsored by the American Farm Bureau Federation) was that it did not seem to warrant the

EPA's conclusions. On the other hand, it has been shown in various studies of certain species of test animals that TCDD can cause cleft palate, kidney abnormalities, fetal mortality, and reduced fetal weight.

Even though sophisticated techniques were used in these studies, traces of TCDD were found in a few cattle in some areas where 2,4,5-T had been used, while in a number of other studies no traces of TCDD have been detected in animals. Some scientists maintain that more sensitive instruments might have detected TCDD and that even extremely low concentrations are hazardous. Furthermore, people do not yet agree on all aspects of interpreting the dose level responses, in other words, whether or not there is a level of this chemical below which no effect will occur. The question also remains whether or not 2,4,5-T and/or its contaminant TCDD might pose a threat of cancer, birth defects, or abortion to human populations. Certain groups vehemently continue to oppose any use of 2,4,5-T no matter how little TCDD it contains. Currently, the recommended maximum level of TCDD in 2,4,5-T is 0.10 ppm. Dow Chemical Company reports that more recent technology has made it possible to reduce the level of TCDD in the 2,4,5-T it produces to less than 0.02 ppm.

When the use of a substance such as 2,4,5-T is suspended, hearings can be requested on the cancellation of that chemical's registration. In July of 1979 the EPA issued notice of an intent to hold a hearing to determine whether or not the remaining nonsuspended registered uses of the pesticide (on rangeland and rice, and for noncrop uses) should be cancelled. In September of 1979 the Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel recommended further reductions of the level of TCDD in 2,4,5-T, obtaining more long-range data on exposure, and evaluation of information from animal studies that had not yet been completed. The Scientific Advisory Panel also said it "could find no evidence of an immediate or substantial hazard to human health or to the environment associated with the use of 2,4,5-T" on rice, rangeland, orchards, sugarcane, and in certain noncrop uses. It therefore recommended that the EPA not hold a meeting at that time to resolve the fate of the nonsuspended uses.

Nevertheless, the EPA is currently conducting hearings on both the suspended and nonsuspended uses of 2,4,5-T. Evidence is being collected from the results of animal and epidemiological studies, and new facts are emerging. For example, it has been reported that many of the 75 chlorinated dioxins may be generated from sources other than 2,4,5-T, such as fuels and other common materials. The full significance of this finding is still not understood, but when an environment exposed to 2,4,5-T is examined for TCDD, measurable amounts of

TCDD may or may not be found. However, when there is a search for TCDD and other chlorodioxins in an environment exposed to large sources of combustion by-products, measurable amounts of TCDD are found. More information is being assembled as various studies are completed. For example, scientists were recently unable to detect dioxin in mother's milk in areas where 2,4,5-T had been used. No residues were detected using the most modern scientific equipment—equipment capable of measuring residues of 1 to 4 ppt. It is not known whether any dioxin is present below this limit of detection. In December of 1980 the Advisory Commission on Pesticides for the United Kingdom issued its most current report concerning the safety of 2,4,5-T. That report concluded "that 2,4,5-T herbicides can safely be used in the United Kingdom in the recommended way and for the recommended purposes"; however, the committee also expressed its intention to "continue to examine any soundly based new evidence or information."

The hearings on the 2,4,5-T issue in the United States are still underway. Because of the many complex issues involved, it probably will be some time before they are concluded and important risk/benefit questions about 2,4,5-T and a host of other chemical pesticides are resolved.

ALTERNATIVE METHODS OF PEST CONTROL

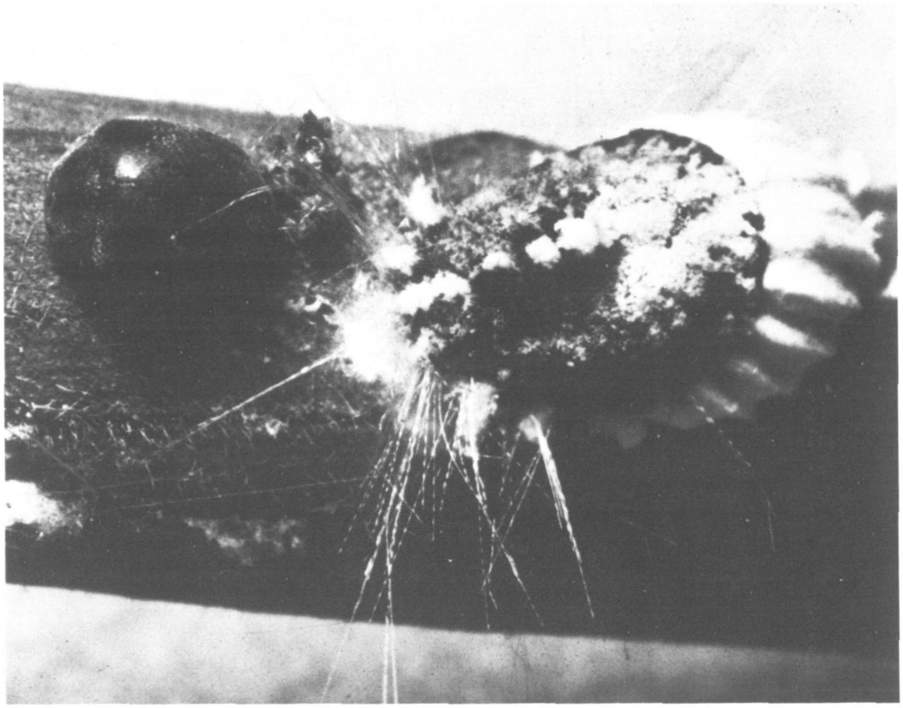
The public concern mobilized by the publication of *Silent Spring* and the report by President Kennedy's Science Advisory Committee has brought about a serious search for alternatives to the heavy use of chemical pesticides.

To achieve the goal of reducing the use of chemical pesticides, the concept of Integrated Pest Management has become a research priority for industries, universities, and government. A number of new, non-chemical methods of pest control have come under study and have progressed to field testing with varying degrees of success.

Biological Controls

The greatest efforts in biological control have been directed at insects and mites. Hundreds of years ago the Chinese used ants to control insect pests on citrus plants. But it was not until around 1890 that a major biological control effort was undertaken in the United States.

At that time the cottony cushion scale, *Icerya purchasi*, nearly destroyed the beginnings of citrus farming in California. This scale was found in Australia but was not considered a pest there because of a natural enemy—the predatory vedalia beetle, *Rodolia cardinalis*. Within a year after the vedalia beetle was imported and established in the California citrus groves, the scale problem was under control.



Around 1890 the cottony cushion scale (right), threatened the success of citrus farming in California. Its natural enemy, the predatory vedalia beetle (left) was used to control this pest in the first major biological control effort that was undertaken in the United States.
APHIS-USDA

The suppression of this pest continued for more than 60 years until DDT was used in some groves to combat other pests. Unfortunately the DDT killed the beetles, too, and there was a widespread recurrence of cottony cushion scale. Control was achieved through the reintroduction of the vedalia and by adjusting the spray treatments.

Insect diseases can be important natural regulators of insect populations, but it is difficult to predict where and when an outbreak of a pathogen will occur. There have been some cases of successfully introducing an insect pathogen to control certain insect pests. One notable such program was designed to control the Japanese beetle (*Popillia japonica*), an insect that injures shade trees, ornamentals, lawns, and various food crops. Estimates of the average damage caused each year in the United States by this beetle range up to $20 million. The beetle lays its eggs in the soil, and the eggs hatch into larvae that eat plant roots before they enter their inactive pupal stage. The adult beetles feed on

Japanese beetles (right) attack flowers, foliage, and fruit of more than 275 different plants. In their grub stage, the beetles are particularly destructive to grass.
USDA

In the photograph below a dust containing bacterial spores from Bacillus popilliae *is being applied to the grounds of National Airport in Washington, D.C. to prevent spreading of the Japanese beetle to distant locations. The bacteria are an ideal weapon because they attack only Japanese beetles and a few other insects in the same family. The disease will not infect other insects, earthworms, birds, warm-blooded animals, human beings, or plants.*
APHIS-USDA

several types of ripe fruits, especially peaches. It was found that a bacterium, *Bacillus popilliae*, can infect the beetle larvae. This pathogen is propagated by injecting the live larvae of the Japanese beetle. Later, the spores of the bacterium are removed from the larvae, dried, and mixed with talc to achieve the desired concentration of spores. The "spore dust" is then spread on the soil to attack the larval beetles. The infected larvae further inoculate the soil and add to the concentration of spores to the point that the beetle population is controlled. In the eastern United States, this method of control is one of the most important biological factors in regulating the number of Japanese beetles.

Another commercially successful effort to control insects with bacteria has been the use of a bacterium named *Bacillus thuringiensis* to control pests such as the cabbage looper, *Trichoplusia ni*, and the corn earworm, *Heliothis zea*. *B. thuringiensis*, has been used commercially for about 20 years, but it does have its drawbacks. It and other microbial controls are highly specific; chemical insecticides must, therefore, be used to control the aphids, mites, and other pests in the same field. The bacteria do not persist in the field, so applications must be made frequently, as with conventional insecticides. Furthermore, bacterial insecticides do not produce the quick knockout of pests that farmers like to see.

Another successful use of biological methods is in the control of the prickly pear cactus, *Opuntia stricta*, on about 60 million acres in Australia. This unwanted plant has been controlled by introducing moths, *Cactoblastis cactorum*, which feed on it. Generally, however, the biological control of a single species of weed is not practical because other weeds quickly take over when the weed causing the primary problem is gone.

The successes of biological controls have been limited. Unfortunately, of the approximately 85,000 species of insects living in the United States and Canada, only about 60 species are reported to have been partially, substantially, or completely controlled by managed biological intervention.

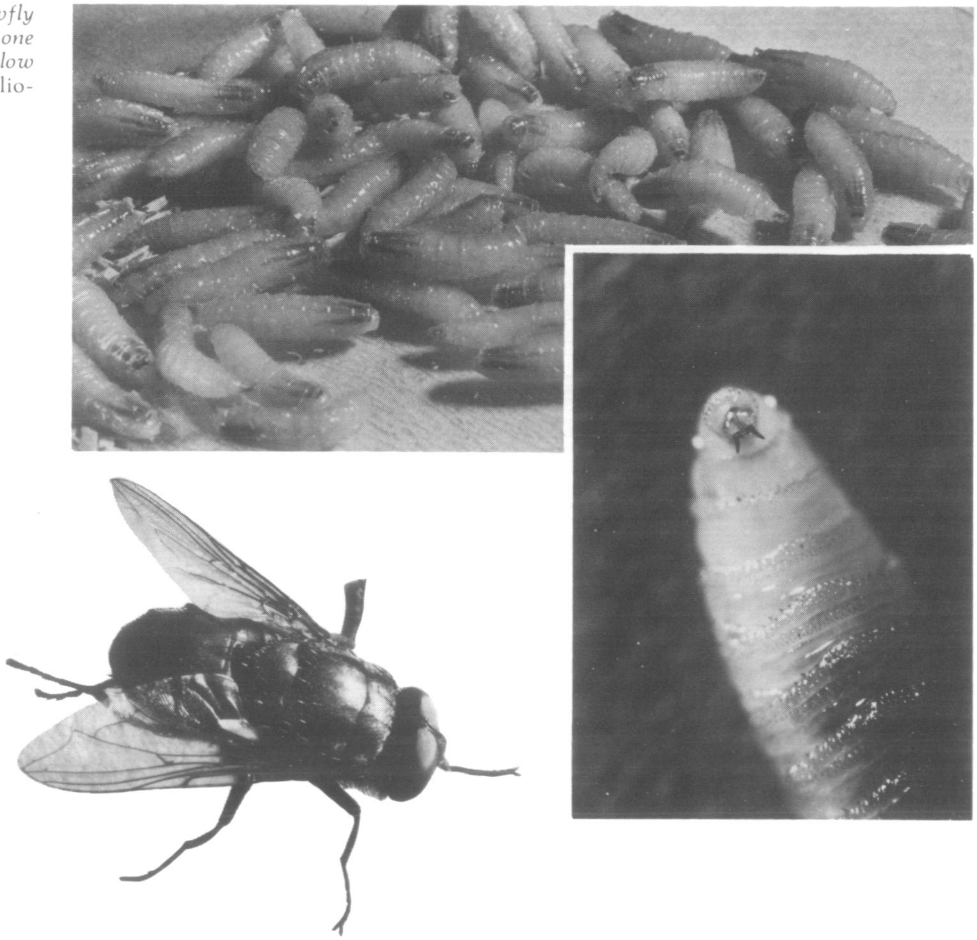
Sterile-Male Technique

Sterilization of male insects by radiation as a means of pest control or eradication was conceived by E. F. Knipling some 40 years ago. His theories were first put into practice in 1955 to eradicate the screwworm, *Cochliomyia hominivorax*, in Curacao.

Screwworms are the maggots of blowflies. The maggots eat the flesh of living cattle, and the wounds they cause may kill the cattle or severely damage the quality of the cowhide. In the sterile-male technique, male insects are raised in an insectary where they are treated with radiation to make it impossible for them to reproduce. These sterile males are then released to mate with females who later lay eggs that develop improperly or not at all. Releasing sterile males has been quite successful in almost eliminating the population of screwworms in the southeastern United States. The techniques involved in the successful control of the screwworm have rarely been duplicated for other pests during the past 25 years.

Researchers are also exploring the use of chemical sterilization, but at this time all of the substances that are effective chemosterilants have harmful effects on higher animals and human beings and are hazardous even under carefully supervised conditions. Unfortunately, it appears that the sterile-male approach to insect control (although scientifically interesting and feasible for certain insects like mosquitoes and fruit flies) is not apt to have much broad impact on reducing chemical insecticide usage in the near future.

At right is a group of blowfly maggots (screwworms) with one enlarged to show detail. Below is an adult blowfly (Cochliomyia hominivorax).
APHIS-USDA



Hormones and Pheromones

Two of the most highly publicized methods of "natural" control are the use of insect hormones and pheromones. The hormones can interfere with the natural physiological processes of a pest, and the pheromones can affect a pest's behavior.

A number of insect hormones have been identified. The hormones are used to kill by interfering with, confusing, overstimulating, or inhibiting the insect's normal physiological functions. They may act on the insect's body chemistry and may affect its gut activity, blood sugar level, or fertility.

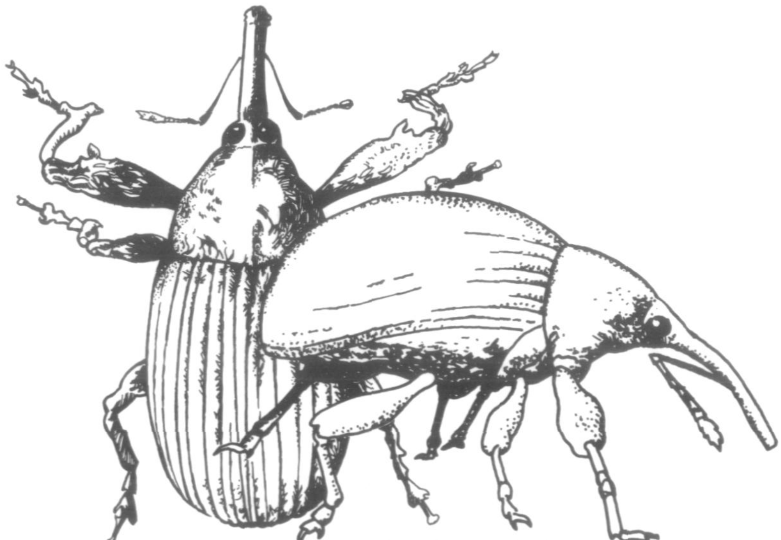
One method being used to control blowflys involves breeding and hatching them in a laboratory where the males are sterilized by radiation. The resulting sterile male blowflies are then released in an infested area to mate unsuccessfully, thereby reducing the blowfly population in that particular area.
APHIS-USDA

Pheromones are like hormones in that they are glandular in origin and act in minute amounts. The big difference is that while hormones are secreted internally and act on the pest's body chemistry, pheromones are secreted externally by one creature and act on another, usually of the same species. Pheromones exercise behavorial control over colonial insects through their odors which may cause clustering, indicate alarm, or blaze an olefactory trail for others to follow. Probably their best known function, however, is as sex attractants. Pheromones of important insect pests are being isolated and identified, and some are available commercially. Traps containing sex pheromones are used to attract insects so their populations can be monitored. Another use of pheromones in the management of pests is to suppress mating by confusing the male insects. This procedure is still experimental; but it rests on the theory that if many sources of synthetic sex pheromones are introduced into the native environment of the insect, there will be a greatly reduced possibility of the male or female insect finding a member of the opposite sex.

Methoprene, produced by Zoecon Corporation, is one of the few hormonal pesticides now available. It is used as a flood-water mosquito larvicide and also for the control of hornflies (*Siphona irritans*). Although use of hormonal pesticides and control by pheromones are not widespread, these two areas of research are promising. A considerable amount of thought and work are being devoted to making these methods more useful.

A trap such as the one shown (below right) can be used to lure unsuspecting boll weevils with a chemical that duplicates their natural scent.

APHIS-USDA



BOLL WEEVILS

Crop Resistance

A successful method of nonchemical control of destructive insects has been the selection and development of resistant crop varieties. Almost everyone is aware that the development of hybrid seed corn revolutionized farming in the United States so much that yields have nearly tripled in the past 50 years. However, few people realize that research to improve seed corn is still very active. One major goal of this research is to develop hybrid varieties that are resistant to pests and diseases. As an example, the research center of Pioneer Hi-Bred International Inc. grows corn borer moths, *Ostrinia nubilalis*, the larvae of which can easily destroy corn crops. The company grows the moths to obtain more than 60 million eggs a year. The newly hatched larvae from these eggs are used to test different strains of corn for their resistance to this pest.

The search for pest-resistant varieties must be a continuing process because living organisms can acquire resistance or immunity to nonchemical as well as to chemical control measures. When pressure is applied that comes close to exterminating an organism, the threatened organism may be able to adapt to overcome the threat, particularly when long periods of time are involved. The process of natural selection of insect-resistant and disease-resistant strains of crops is going on continuously, and research such as that described with corn is needed to keep the farmer ahead of the pests.

Disadvantages

One major problem facing the further development of nonchemical methods of pest control is their specificity. Because these methods usually are effective against only one kind of pest, crops still require chemical or other treatments to handle threats from other species of pests. The second important issue for nonchemical (and chemical) control methods is the possibility that health or environmental hazards may be introduced which have effects at least as severe as those methods they are intended to replace. This is particularly true in the cases of hormonal controls and chemical sterilization techniques.

A third major drawback in the use of nonchemical measures is the cost of developing and marketing them. Because of the specificity of these measures, a feature that many environmentalists consider desirable, any single product will have a limited market. So, even if a product could capture the entire market for the control of a particular species of pest, its limited demand might not justify the costs of development.

INTEGRATED PEST MANAGEMENT

The most widespread attempt to integrate nonchemical controls with chemical controls is Integrated Pest Management (IPM). Researchers using the biological, cultural, and chemical control phases of Integrated Pest Management programs have provided valuable knowledge needed to attack a pest at its most susceptible stage and to determine the circumstances under which the host is most vulnerable.

Few would argue about either the desirability or the potential of combining multiple methods of pest control into an effective Integrated Pest Management system. But the lack of research information on the basic biology of pests and crops, the lack of established economic thresholds, and the primitive state of predictive modeling and agroecosystem analyses are serious deterrents to the growth of IPM.

Although according to the 1979 *Report on Pest Management Strategies* of the Office of Technology Assessment

> IPM appears to be the most promising crop protection strategy for the next 15 years, [there are] technological and administrative obstacles that impede the development and implementation of IPM.
>
> The technological obstacles lie primarily in the areas of basic knowledge, delivery systems, and personnel. An inadequate base of knowledge in the biology, bionomics, and interactions of crop pests seriously limits the range of control tactics available for integrating pest management into a total crop production system.

OTA has concluded that, with increased use of IPM, chemical pesticide use is expected to decrease on all crops considered in the 1979 report; suprisingly, however, it has also concluded that "the amount of reduction is speculative and may not be nearly as significant as is often assumed by some persons. It is more certain that the pesticides that are applied will be used efficiently and effectively."

Because chemical pesticides will continue to be heavily used in the foreseeable future, it is important to discover for each crop and each pest how to select the most specific chemical available, the most suitable formulation and application techniques, the optimal dosage, and the proper time of application. Some such knowledge is part of the experience background of any successful farmer. But to get the best results, it is necessary to have information about trends in pest populations obtained by areawide surveillance and an understanding of the relationships between pest populations, pest damage, crop production, and profits. For example, in some cases, while a pest-control technique may reduce damage by pests, it may also decrease overall production. In other cases, the cost of pest control may exceed the benefits in increased crop production, making it more profitable to accept reduced production and to avoid the expense of certain methods.

Convincing skeptical growers that Integrated Pest Management is a viable alternative to their present pest-control practices is not an easy job. Many growers have confidence in what they are doing now; many, already faced with serious threats from pests, are hesitant to abandon current control methods for the uncertain services and benefits offered in an IPM program. And in many cases farmers do not see any economic benefits from IPM. Because of the complexity of IPM strategies, a great deal more research is needed to develop, implement, and improve them.

PEST CONTROL AND SOCIETY

Measuring the social impact of pest control is difficult. The reasons for and results of its use are wide-ranging and complex. One of the more obvious reasons for controlling pests is to enlarge crop production. If a reduction in the number of pests permits farmers to increase their harvest, the farmers may realize greater profits because their unit costs are reduced, and consumers may pay less for food because it is in greater supply. Past absences of or failures in pest-control programs have set the stage for such tragedies as a plague of locusts devastating hundreds of square miles of cropland, causing starvation, and uncontrolled insect populations transmitting disease to millions of people, causing epidemics of typhus, malaria, plague, or encephalitis.

There are other, perhaps less-compelling, reasons for using pesticides. Farmers may find it necessary to apply pesticides to reduce the aphid population on tomatoes to meet the strict tolerances for the number of parts of insects permitted in catsup, even though the average consumer may not be aware of any contamination of catsup. Cosmetic considerations sometimes mandate the use of pesticides. For instance, mite damage may affect the appearance of an orange while having no impact on its taste or nutritional quality. An orange's appearance, however, may affect its marketability—a strong incentive for a farmer to control the mites.

Pest-Borne Diseases

One aspect of pest control that is often ignored in the continuing debate over its need is the contribution of chemical pesticides to world health.

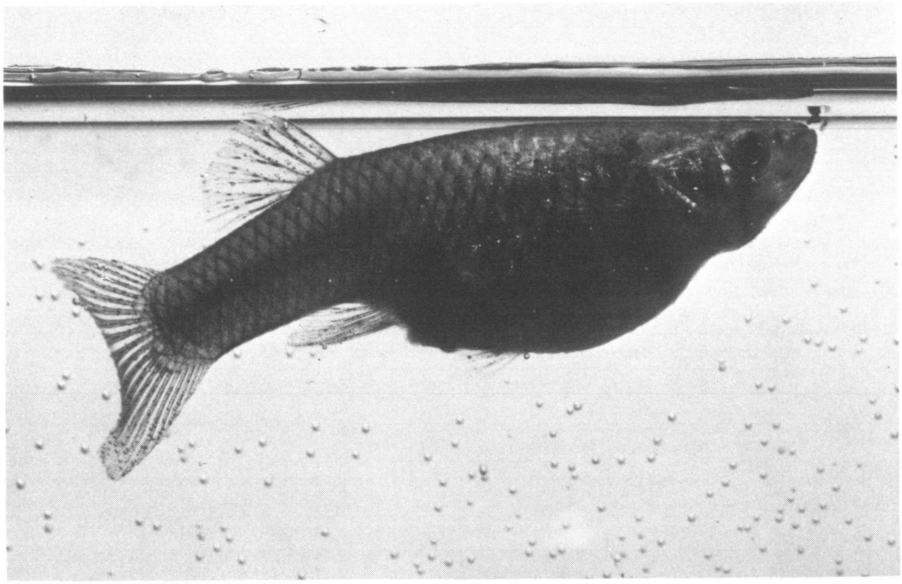
Prior to 1939 epidemic typhus killed more people during all the wars in history than all the spears, arrows, bombs, and bullets combined. One type of typhus is transmitted by body lice, a pest that thrives on the unsanitary conditions associated with extended battle. War and this disease seem to go together.

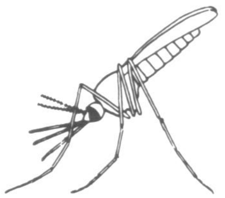
Early in World War II, long after Paul Müller discovered that DDT was effective in controlling a wide range of insects, including lice, the allied forces put DDT into mass production. In 1944 DDT was applied directly to thousands of soldiers, refugees, and prisoners to combat body lice, thereby helping to avoid the spread of a disease that killed more than 2 million people during World War I and its aftermath.

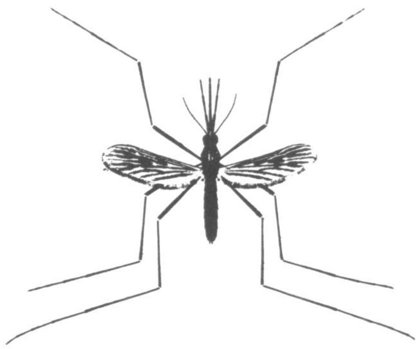
Malaria, probably the greatest killer and debilitator in the world, is carried by several species of the *Anopheles* mosquito. Prior to the widespread use of DDT, control techniques included spreading oil on the surface of water to prevent mosquito larvae from getting air necessary for their survival, water management, and using pesticides such as Paris green and pyrethrum spray.

Through the use of DDT and other insecticides, the elimination of potential breeding places, efforts to improve the standard of living, and the discovery of better medical treatments, the picture gradually changed until the disease was no longer a problem in the United States. Nevertheless, malaria remained a problem in other parts of the world. It is estimated that in the late 1940s and early 1950s there were about 300 million cases annually, with at least one death occuring from the disease every 10 seconds. In 1956 the World Health Organization mounted an international malaria-eradication effort using DDT. Although the program was very successful in many countries, it is estimated that the disease still causes 1 million deaths (mostly in Africa), and there are 120 million cases per year. On the other hand, the persistence of DDT and the harmful effects caused by its accumulation in the food chain contributed to its being banned for use in the United States.

Other serious diseases are also borne by mosquitoes. Yellow fever is still a major threat in Africa although it has been eradicated in most urban areas of the New World. Filariasis, a disease caused by a parasite nematode worm and transmitted by mosquitoes, infects an estimated 250 million people. Mosquitoes are the vectors of more than 80 viruses







Gambusia affinis (above right) is a small fish which is a natural predator of the larval and pupal stages of mosquitoes of the genus Anopheles. Gambusia (right) are being liberated into a pond to help control the mosquito population.

USDA/Center for Disease Control

that attack human beings and cause diseases, such as o'nyong-nyong fever, dengue fever, Chikungunga fever, and several types of encephalitis.

Encephalitis, a disease present in the United States, is transmitted to people by birds and mosquitoes. In most years there is little risk to people, but when mosquitoes become more abundant—after heavy rains for instance—the risk increases. Researchers are developing a number of nonchemical alternatives to prevent outbreaks of encephalitis in this country. Many of these techniques could have applications in other parts of the world to control mosquito-borne diseases.

There is evidence that the mosquito now is exhibiting incredible adaptability. Southern Asia and parts of Central and South America are experiencing a resurgence of malaria. Mosquitoes are developing a high resistance to household sprays and other methods used to control them.

In addition to insecticidal sprays there are other approaches to controlling mosquitoes. Just as some people have hereditary resistance to certain diseases, mosquitoes have been found that resist laboratory efforts to infect them with encephalitis virus. Such mosquitoes may be released in hopes of eventually replacing those which carry the disease. The use of natural predators (for example, fish that eat larvae) is another control method which is effective in certain locations.

Biologists first thought that the variations in the number of mosquito larvae from rice field to rice field in California resulted from toxins from a blue-green algae. Recent research, however, has shown that when mosquito larvae brush against certain species of flatworms in the genus *Mesostoma* the larvae become paralyzed and die. Studies are now underway to find out why certain rice fields support flatworms and others do not. These worms may provide another biological approach to mosquito control. Parasitic roundworms that can kill larvae may also prove to be the mosquito's nemesis.

These are just a few methods that have potential for controlling encephalitis-carrying mosquitoes in the United States. Many of these methods may have application elsewhere for controlling pests which transmit other diseases; enormous health benefits could ensue if the techniques are successful.

Risks and Benefits

When spray operators become ill or die as a result of spraying or spilling pesticides on themselves, the adverse effects of pesticides are immediately evident. In other instances determining that a death has been caused by a pesticide often requires a little more detective work. Accidents caused by careless storage of pesticides around small children and even suicides show up in the statistics on pesticide poisoning.

Immediate and fatal effects are fairly easily documented. Despite the vast increase in the availability and use of pesticides, since 1956 direct deaths attributed to poisoning by agricultural chemicals, including pesticides, dropped from 152 in that year to 34 in 1977. (In 1977 the total number of lethal poisonings from all substances in the United States was 4,970.) During this same period the total population and the number of accidental deaths from all other poisonings approximately doubled.

Pesticides may cause unwanted damage to crops and animals which can occur at the site of application or many miles downwind or downstream from the point of application. Some other adverse effects are harder to identify. In general, it is rare that the cause of an individual case of cancer can be identified reliably. Epidemiological studies of population groups may associate a rise in rate of cancer with a common exposure to a chemical, but such detectable outbreaks are usually in relatively small, immobile, and homogeneous populations of heavily exposed individuals. No such outbreak of cancer has been reliably related to a pesticide—although the possibility of such a relationship cannot be dismissed. The United States Environmental Protection Agency has established a Cancer Assessment Group with the responsibility of evaluating evidence that certain substances are or are not carcinogenic. Their deliberations are principally based on the results of experiments on laboratory animals and findings in epidemiological studies.

Problems also exist in assessing the more serious long-term or delayed effects of pesticides on wildlife and on the environment in general. As already mentioned, certain pesticides persist in the environment and may accumulate in the food chain. Few doubt that the potential exists for serious, long-standing ecological damage, but such damage is difficult to demonstrate in many specific cases and is, therefore, almost impossible to quantify. It is not uncommon for a pesticide to kill beneficial parasites and predators of pests as well as the pest itself, thus resulting in a serious infestation of the same or a different pest at a later time. But this effect is also difficult to measure. Biological control methods may also suppress the population of one pest, thereby, in turn, allowing the population of another pest to develop into a serious problem.

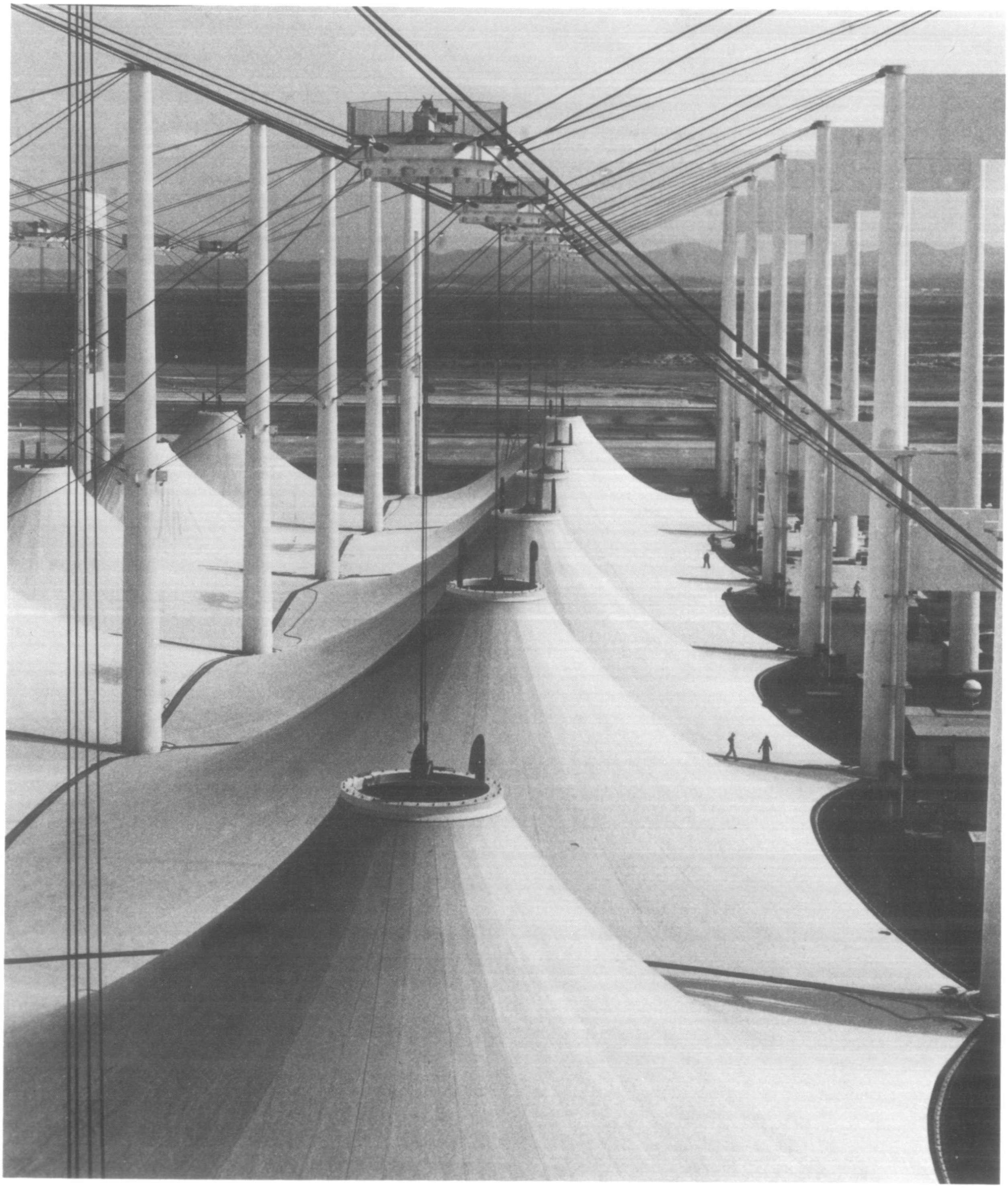
Effects of pest control measures are hard to classify as entirely beneficial or adverse. It is easy, for example, to understand why the Mormons in Utah erected a monument to the seagulls that saved their crops from a devastating horde of insects. It is somewhat more surprising to learn that a monument was erected to the boll weevil in Enterprise, Alabama, because that insect forced farmers to diversify their crops, thus benefiting the entire state.

Few experts in the field of pest control believe that it is possible for the United States to maintain its current levels of agricultural productivity, let alone to increase production for expanding world needs, without some use of chemical pesticides. Similarly, acceptable freedom from pest-spread diseases, pest annoyance, and the destruction of crops and forest products by pests cannot be achieved without some use of chemical pesticides. On the other hand, the tremendous surge in chemical technology several decades ago, especially the rapid conquest of serious pest problems during World War II, resulted in an unjustified confidence on the part of some people that chemicals could solve all pest problems. Scientists were less sanguine about having found the ultimate solution to pest control than the users were, and the scientists have been proven right. Today the need to reduce damage, health hazards, and annoyance from pests is still present.

Accomplishing effective pest control will require more research: the hazards of pesticides need to be reduced; the ecological interactions among people, pests, and environment need to be better understood; the ability of those concerned with pest control to communicate among themselves needs to be increased; and the technology for dealing with the complexities of the pest-control equation needs to be improved.

Answering questions involving the risks and benefits of pesticides in today's society is the responsibility of legislators, regulatory agencies, county and state research and advisory services, agricultural chemical companies, farmers, and finally, individuals. All have a vital stake in the continued successful management of pests.

The present political reality in the United States and in other parts of the world is that consumer and other interested groups object to the use of a technology unless they have been persuaded that the benefits of the technology are great enough to justify the risks that are involved. Within the last decade these groups have awakened to the fact that pressure from social forces can result in the banning of hazardous pesticides. It is likely that this monitoring process will continue; it is clear, therefore, that the bases of assessing risks, benefits, costs, and the ways in which they are distributed to various segments of society must be better understood. To this end, both experts in the field and interested segments of the population must learn how to communicate more effectively with each other. If mutual respect among the various segments grows, it should be possible for adversaries to understand and agree on the nature of the risks, benefits, costs, and alternatives, even if they attribute different values to each.



Synthetic Fibers

About 30 years ago the term "wash-and-wear" was introduced to the American public. Since then it has become virtually synonymous in the American mind with any fabric made entirely or in part of fibers created in the laboratory. The easy care that is characteristic of synthetic fibers is already legendary; over the years their durability has also proved undisputed.

When the Haj terminal at the new Jiddah International Airport is completed, it will be covered by 5.5 million square feet of Teflon®-coated fiberglass fabric. The translucent fabric will form 210 tent-like units spanning 105 acres—an area equivalent to that of 80 football fields. The terminal's principal purpose is to serve Muslims making their once-in-a-life-time holy pilgrimage to Mecca. Six hundred thousand pilgrims are expected to use the facility in 1981. This building material was chosen because of its high strength, resistance to weather, and long life.*

Owens-Corning Fiberglas

For thousands of years, fabric had been woven from natural fibers, such as silk and wool from animals and cotton and flax (linen) from plants. Not until the twentieth century did scientists and engineers successfully reach beyond the range of fibers provided by nature. Research made it possible to use feedstock chemicals to produce synthetic fibers, and powerful new technologies provided the means to reform, shape, blend, coat, and treat the fibers.

Today more than one-third of the fiber used in the world and about three-fourths of that used by mills in the United States are synthetic. Synthetic fibers are everywhere—providing clothing, shelter, and household furnishings, improving long-range communications, and helping make more time available for leisure and for work. Collectively, in their many applications, synthetic fibers make the modern world more habitable.

THE VERSATILE FIBERS

Today, fiber chemists can produce and tailor synthetic fibers to meet a variety of special needs. Synthetic fibers also can be produced more rapidly and in some cases more economically than natural ones, especially in recent years. Polyester staple, for example, is usually cheaper than cotton, and nylon and polyester filament yarns are cheaper than wool and silk—primarily because of the differences in the amounts of labor required to produce natural and synthetic fibers.

The textile manufacturing process tends to become increasingly labor-intensive as the raw product goes from fiber to yarn to textile to finished product. Making the fibers, dyeing them, and finishing them represent less than 10 percent of the retail cost of a garment, while spinning, weaving, design, and manufacture account for more than 50 percent of the cost. Making cloth from synthetic fibers requires less labor than making cloth from natural fibers. Because of this saving in labor costs, the average family in the United States now spends a much smaller portion of its income on clothing than it did 20 years ago.

Synthetic fibers have made dramatic changes in clothing possible. Today's comfortable, fitted, and functional styles are possible only because of synthetic fibers. Synthetics have even sparked a number of fashion revolutions. Textured nylon, for example, led to pantyhose, which in turn helped popularize the miniskirt. Modern styles of lingerie, swim, and other sportswear owe their existence to synthetic fibers.

Synthetic fibers also mean that not only the wealthy but also a vast spectrum of Americans can enjoy the latest trends in home decoration. The fibers have made possible a great variety of inexpensive, easily maintained, durable, and colorful fabrics for upholstery, draperies, and linens. In fact, bulked, continuous filament yarns have created a carpeted America.

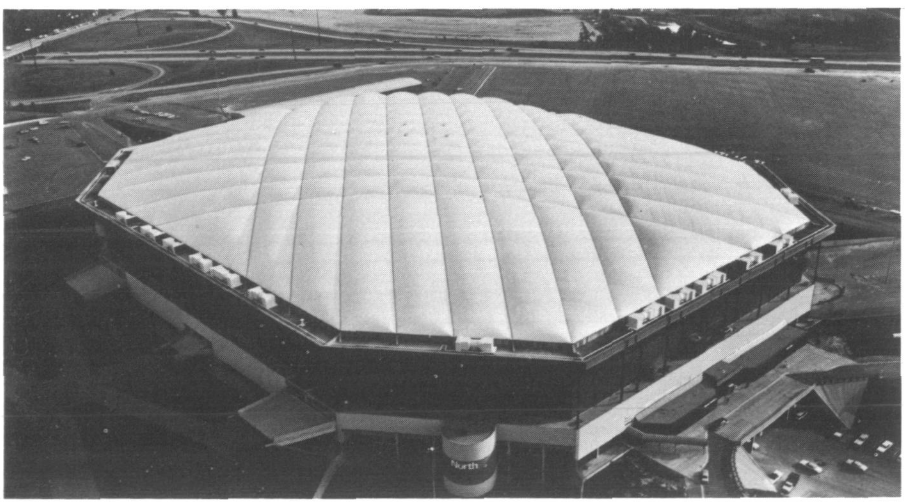
Because synthetics are so easy to maintain, the need for mending has decreased; and although people now wear more knits than they did a generation ago, darning is almost a forgotten art. In the past, an average family's ironing may have taken more than 1 day a week, and this was for approximately one-half or less of the clothing and furnishings that Americans own today. The important effect of these time and labor savings is that people have more time which can be used in other ways. This time can be used to add to the total work force, for recreation, or simply for relaxation.

By cutting the cost of recreation equipment, synthetic fibers have also helped diversify and multiply leisure-time activities. The private boating boom is largely a result of research that made possible the development of synthetic fiber and polymer composites that can be molded into hulls reinforced with glass fibers. Many of the low-cost, sturdy, lightweight tents, sleeping bags, and other hiking and camping equipment more resistant than ever to extremes of heat and cold, wind, rain, and abrasion could not exist without synthetic fibers. In skis, golf club shafts, tennis rackets, tent poles, vaulting poles, and skate boards, where strength-to-weight and flexibility-to-weight ratios are important, fiber-reinforced plastics excel.

The construction industry is using synthetic fibers to cover some long, clear spans. Sport fans know that the Silverdome Stadium in Pontiac, Michigan, has an inflated roof woven of glass fiber. A Teflon® coating protects the roof against extremes of weather, yet the roof remains translucent. The flame-resistant qualities of glass fiber and its high strength-to-weight ratio were important factors in the choice of these materials. When the cost and time to erect this enclosed stadium are compared with the cost and time needed to build similar structures in recent years, the economic advantage of this inflated roof over conventional ones is easily demonstrated.

Inflated structures also have permitted some large-scale construction projects to proceed virtually all winter long in even the coldest climates. Such structures also are being used in hydroponic gardening and farming to make plant growth independent of the weather. Experts predict that the ability to erect these structures quickly and inexpensively over large areas will have increasing application in construction, agriculture, and industries in which pollutants must be controlled.

Even as they are being further refined, synthetic fibers continue to be used for a variety of special environmental uses. These include hollow-fiber membranes used to purify water by hyperfiltration and filter bags that can remove particulate matter from smoke. Nonwoven fabrics made of olefin, nylon, or polyester are serving in road construction as a textile layer between road bed and surface pavement; they



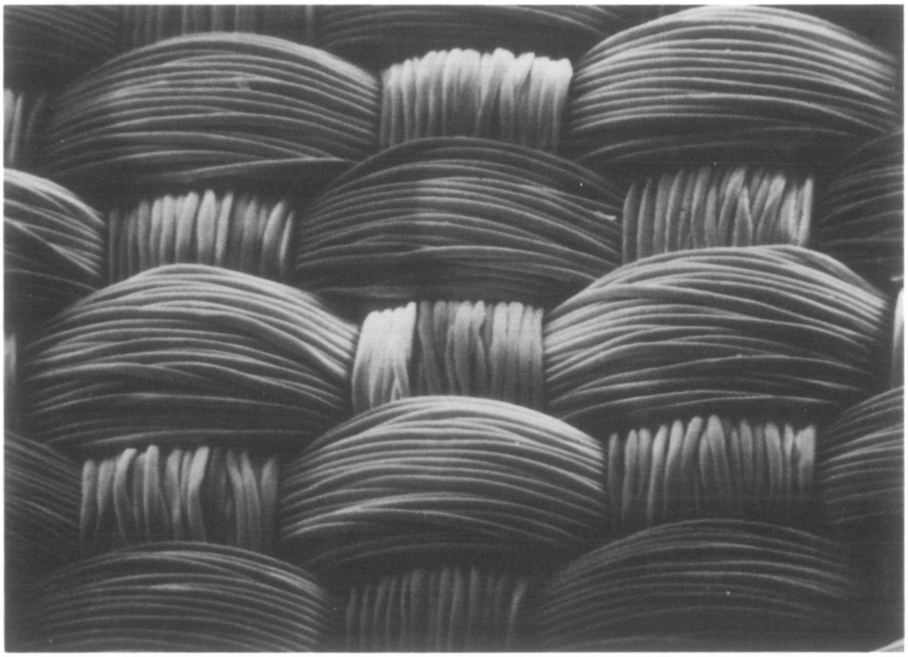
The Silverdome Stadium in Pontiac, Michigan has more than 80,000 seats and covers about 35,000 square meters.

Owens Corning Fiberglas

resist the elements and prevent cracking of the road surface. A less well-known use of synthetic fibers is in a so-called nylon "whale" capable of sucking up thousands of gallons of oil spilled into the ocean.

Not too long after synthetic fibers became available, surgeons started replacing portions of damaged or blocked blood vessels with tubes woven of synthetic fibers. For about 20 years synthetic fibers have formed the soft, woven synthetic fabric that rings an artificial heart valve, permitting it to be sewn into the wall of the heart. Even the suture materials are frequently synthetic fibers.

Synthetic fibers also are used for such nonwoven hospital disposables as bed linens and gowns for professional and nonprofessional personnel. Glass fibers also are used in probes made of special glass or plastic fibers which can transmit light. These make it possible to see inside many of the body cavities and hollow organs, including the abdomen, bladder, chest, and the lungs themselves, without resorting to a major surgical procedure to make a diagnosis or to give treatment. In the last few years, medical scientists have been experimenting with the use of carbon fibers in artificial joints because these fibers are lightweight, very strong, and chemically inert. Indeed, the number and kinds of medical uses for synthetic fibers have increased markedly in the past decade, but these fibers represent only a small portion of the total volume of synthetic fibers used for all purposes because medical devices usually are small.



Human blood vessels may be repaired using grafts made of synthetic fibers such as this Veri Soft woven graft (130x).

Meadox Medicals, Inc.

HOW SYNTHETIC FIBERS ARE MADE

Polymerization is a chemical reaction in which monomers—small chemical compounds with low molecular weight—combine to form a polymer, a more complex molecule with a higher molecular weight and very different chemical and physical properties. Polymerization is basic to the manufacture of synthetic fibers. The process may be visualized as the end-to-end joining of molecules, something like joining the links of a chain. If each link were about ½ inch long, some polymers might be longer than a football field. It is a polymer's high molecular weight and its enormous length-to-width ratio that account for some of its unique properties. Polymers can be molded or shaped; they can be set by heat, chemical processing, radiation, or by a combination of all three processes. Some are more elastic than rubber, and others are so tough they are used to produce helmets and bulletproof vests.

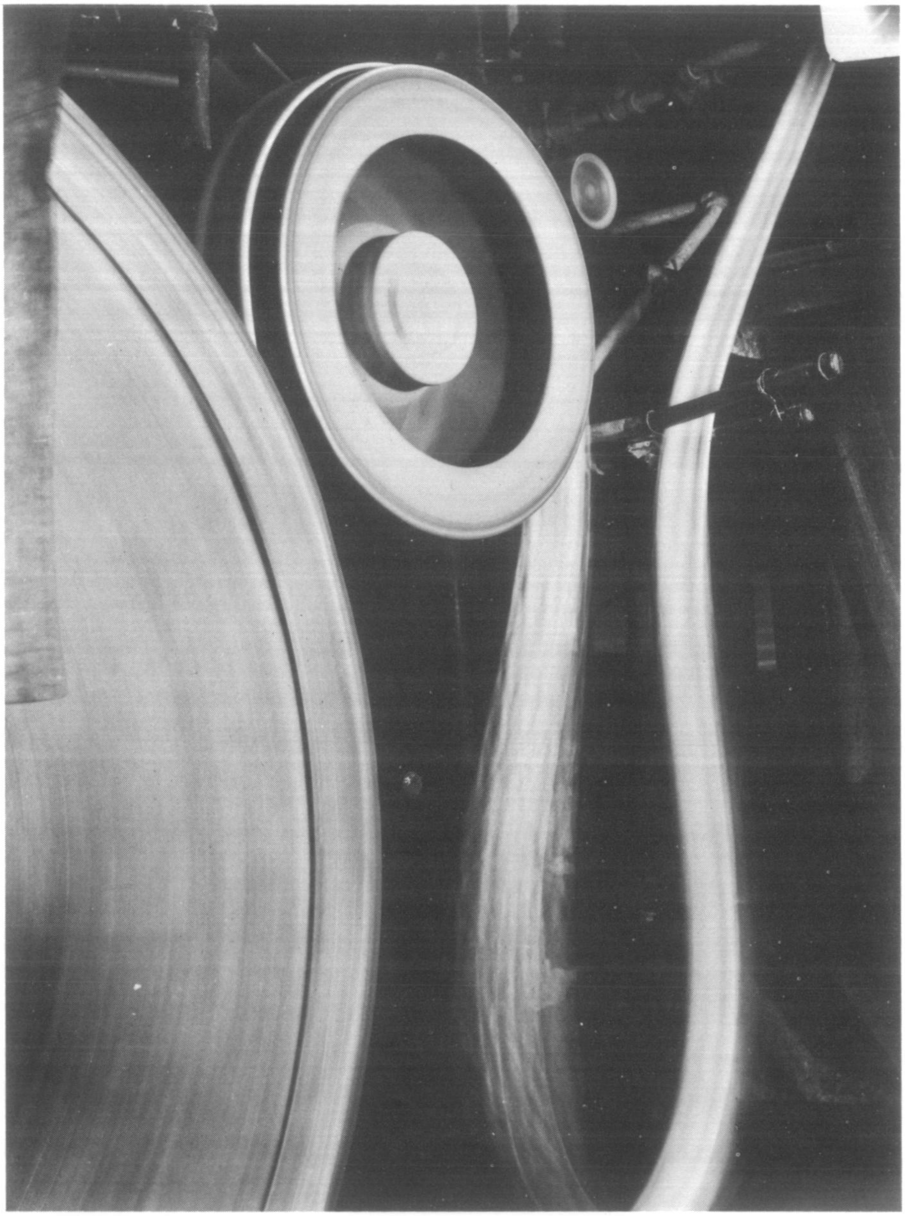
Production of a synthetic fiber begins with liquefaction of a polymer by dissolving it in a solvent or by melting it into a syrupy liquid. It then is pumped through a spinnerette, a nozzle-like device similar to a showerhead containing tiny holes.

As the filaments emerge from the spinnerette, they are hardened or solidified. This process is called spinning. It is spinning somewhat in the manner in which a spider is said to spin and should not be confused with the textile operation that produces yarn.

In the manufacture of synthetic fibers, there are three methods of spinning—wet, dry, and melt. In wet spinning, filaments from the spinnerette pass through a chemical bath where hardening takes place. Rayon is produced by this technique. In the dry method, the extruded filaments dry and harden in warm air. Synthetics processed by this method include modacrylic and acetate. Fibers produced from polymers that are melted before extrusion and hardened by cooling include nylon, polyolefin, and polyester.

A separate process in each of these techniques stretches the fiber after they are extruded from the spinnerette, reducing the diameter of the fibers in proportion to the amount of stretching. This stretching aligns a substantial fraction of the polymer's chain-like molecules parallel to the length of the fiber, greatly increasing tensile strength and stiffness.

Synthetic fibers can be blended with one another or with natural fibers. Two different polymers can be extruded side by side and



This machine was developed during the early days of nylon production. It pressed nylon filaments into a long ribbon, which was then cut, melted, and extruded to form individual nylon fibers of varying thicknesses.

Eleutherian Mills-Hagley Foundation, Inc.

twisted together, or they can be mixed and extruded as a single fiber. Such fibers can give the finished product desirable characteristics, such as antistatic, elastic, or flame retardant properties.

Synthetic fibers can be made in different shapes and thicknesses; they can be textured to add bulk, stretched to conform to desirable shapes, or produced in varying lengths for blending with other fibers. With special additives, basic fiber-forming solutions can produce fibers that are as comfortable as cotton or that look and feel like fur, wool, or silk. Some synthetic fibers can be manufactured into fabrics that are both sheer and strong at the same time.

The various processes available for the manufacture of synthetic fibers and the wide choice of polymers and molecular weights give these fibers their versatility in meeting many different needs.

RESEARCH AND DEVELOPMENT

Rayon (Artificial Silk)

The roots of fiber technology, like the roots of many early technologies, are traceable to Europe. The earliest proposal for making artificial silk is attributed to a versatile English scientist, Robert Hooke. His observations of silk worms and spiders prompted him to boast that he could make a better fiber than the silkworm. In 1664 he suggested that if a suitable liquid could be forced through small openings, it would harden into fibers or artificial silk. Hooke, however, was a man of varied interests; he did not follow through on many of his ideas, and he tended to abandon any slow-moving experiments.

In the 1880s, a Swiss chemist, Georges Audemars, also was interested in making artificial silk. He reasoned that the mulberry tree could provide people, as well as silkworms, with the necessary raw ingredients for making silk. He experimented with the cellulose (the chief constituent of the walls of plant cells) from the inner bark of the mulberry tree to form a suitable liquid, but he lacked a process for extruding the liquid.

Britain's Sir Joseph Swan solved Audemars' dilemma. Swan found that he could produce fibers by forcing liquid through a small opening into a coagulating bath that hardened the material into fibers. But because Swan was interested in making carbon filaments for electric lamps, he never pursued the possibility of making textile fibers. It was Count Hilaire de Chardonnet, a French chemist, who went on to develop a process capable of producing artificial silk, later called rayon. (Strictly speaking, rayon is not usually considered a synthetic fiber because it is made from a naturally occuring polymer, cellulose.) Fabrics were first displayed by de Chardonnet in 1889 at a Paris exhibition; about 2 years later he opened the first commercial rayon plant in France.

In the meantime three English scientists, C. F. Cross, E. V. Bevan, and C. Beadle, were working to improve de Chardonnet's process for manufacturing rayon. They developed what is called the viscose process, the usual way of converting cotton or wood cellulose into rayon.

Two Swiss chemists, Camille and Henri Dreyfus, patented another method, the acetate process, which changes the nature of cellulose by treating it with acetic acid and other chemicals. These two brothers made and marketed low flammability lacquers, plastics, and films.

Most experts agree that fiber technology came to the United States in 1910 when the American Viscose Corporation, now Avtex Fibers, Incorporated, began manufacturing rayon in Marcus Hook, Pennsylvania.

Acetate fiber was first successfully produced in the United States on Christmas Day in 1924 at a plant established by the Dreyfus brothers in Maryland; that plant was the predecessor of today's Celanese Corporation. Viscose and acetate rayons were popular with American shoppers throughout the 1920s.

Nylon

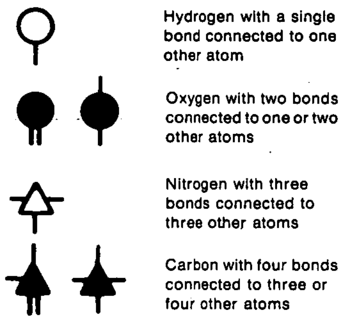
In his *Principles of Polymer Chemistry*, Paul Flory points out that the overriding goal of most chemists around the turn of the century seemed to be the desire to prepare or isolate pure substances; that is, substances made exclusively of a single type of molecule. This objective continued to dominate synthetic chemistry in the 1920s. A chemical discovery was not accepted unless its composition were confirmed through laboratory analyses and the determinations of its proposed molecular weight were shown to conform to the molecular structure.

Although synthetic chemistry produced thousands of different combinations and permutations of atoms, the horizons of research were restricted. Most organic and physical chemists came to believe that every definable substance could be classified in terms of a single discrete molecule that could be represented by a concise formula. Theoretical chemists focused attention on "the molecule," postulating laws for ideal molecules which were visualized as small spheres not much larger than single atoms. According to this concept, the full understanding of the molecule would serve to explain all of the properties of a substance. This perspective, however, overlooked a very important group of substances, the polymers, which cannot be described by conventional molecular formulas and which are the building blocks of natural fibers.

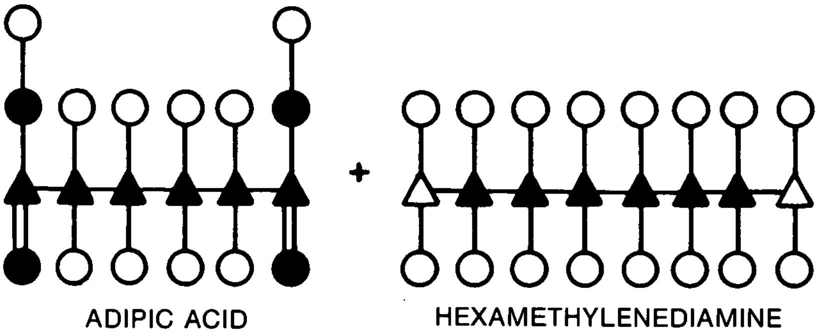
Polymers are much larger than most other chemicals. For example, a simple sugar, such as glucose, has a molecular weight of 180, while cellulose, (a natural polymer) is formed from hundreds of glucose molecules chemically combined end-to-end.

The mysterious physical and chemical nature of polymers took years of investigation to unfold. Until the 1930s several theories about

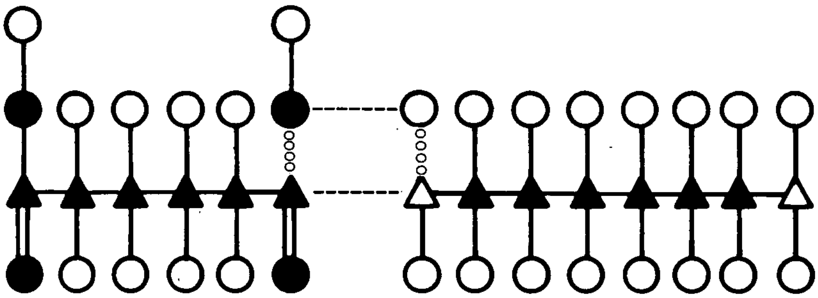
How a Long-Chain Polymer, Nylon-66, is Formed.



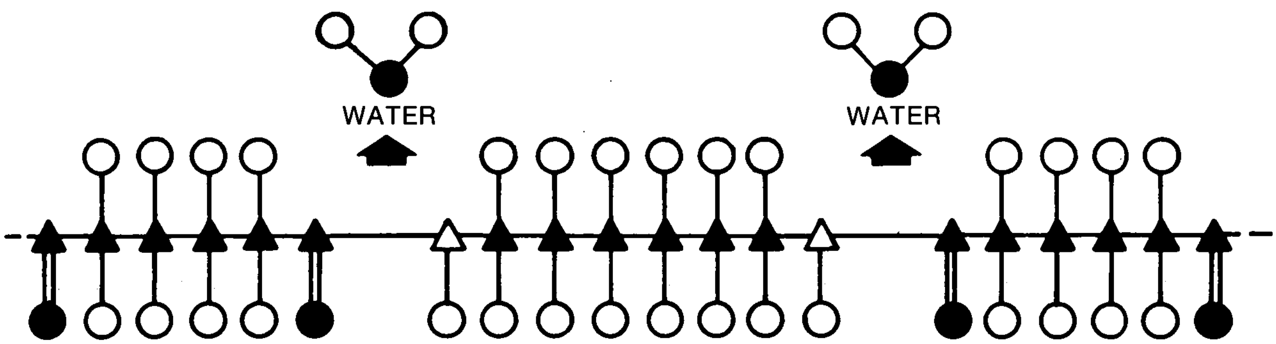
Each type of atom tends to form a certain number of bonds.



One molecule of adipic acid and one molecule of hexamethylenediamine (in which each atom has the desired number of bonded atoms) are brought together.



When the two molecules are brought together and heated, old bonds are broken (ooooo) and new bonds (----) are formed to make a water molecule and a polymer fragment. One end of a former adipic acid molecule is then able to react further with another end of a hexamethylenediamine fragment, thus creating a repeating unit or polymer building block.



This process is repeated many times, fragments are connected, and a long-chain molecule is formed. This particular polymer is nylon-66.

these giant molecules held sway. One was that cellulose was made up of clumps of glucose aggregated in some unknown fashion. Another theory was that relatively simple molecules could combine through strong chemical bonds to form giant molecules—molecules 10 times, or even 1,000 times, larger than the original molecules. The latter was a less popular theory; most scientists investigating the structure of polymers preferred the concept of finite certainty offered by ring formulas to the vagaries of a formula for chains of unidentified lengths.

The divergent theories about the structure of polymers precipitated heated discussions and controversies. Many converts to the giant molecule concept were zealous in their efforts to define the precise structure and properties of polymers. Among the principal proponents of the giant molecule theory were Herman Mark, K. H. Meyer, and H. Staudinger.

Staudinger led the parade of converts to the giant molecule theory, but many thought he had gone too far when he proposed a relatively simple relationship between the viscosity of a solution of a polymer and its molecular weight or size. To disprove the theory, Harold Hibbert and his students at McGill University in Canada undertook an extravagant program of synthesizing ever larger molecules by classical methods to disprove Staudinger's viscosity theory. Their efforts had the opposite result from the one intended. Their research confirmed that polymers are actually chemicals of high molecular weight and that the viscosity of dilute polymer solutions is determined by the polymer's molecular weight. (In 1953 Staudinger received a Nobel prize for his work on giant molecules.)

Once it was determined that natural fibers have unique properties because they are composed of giant molecules, chemists began to synthesize giant molecules from smaller ones and to study their properties. This type of research became one of the major activities at E. I. du Pont de Nemours & Company, Incorporated, in Wilmington, Delaware.

In 1927 the Du Pont Company initiated a research program in organic chemistry. The aim was to gain a better understanding of the chemical processes involved in polymer formation and possibly to find new paths of applied research. A year later a brilliant young chemist, Wallace Carothers, was persuaded to leave Harvard University, where he held the rank of instructor, to direct this work at Du Pont.

From the beginning, members of the Du Pont group were encouraged to select their own projects. Carothers was interested in substances of high molecular weight and in polymerization by condensation. Through his basic research on condensation polymerization, he

In 1927 Wallace Carothers left his instructor's position at Harvard University to head a research group at E. I. du Pont de Nemours & Company, Incorporated, where he initiated and guided the research that eventually led to the discovery and production of nylon.

Eleutherian Mills-Hagley Foundation, Inc.

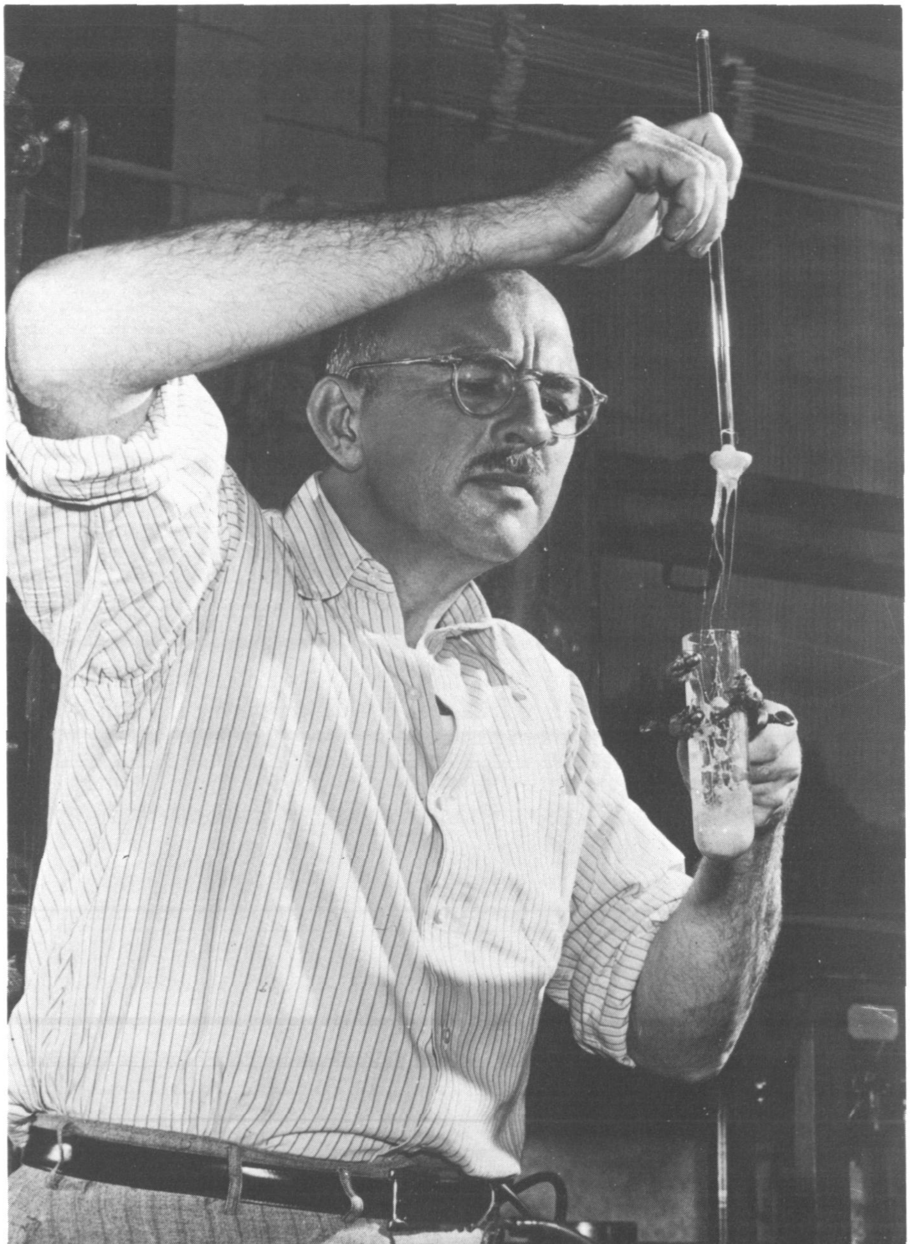
obtained polymers, chemicals characterized by recurring structural units. After investigating the preparations and properties of many such substances, Carothers made a significant advance in their preparation. He used a device called a molecular still that had been set up at Du Pont by Julian Hill, a member of the group. The equipment made it possible to carry polymerization closer to completion. Because the still could eliminate water formed during the condensation reaction, Carothers could obtain linear polymers with much higher molecular weights than ever before. Some of these substances had molecular weights of over 10,000; Carothers called them "superpolymers." (Today polymers with molecular weights of more than 10 million are known.) The polymers which Carothers made in the molecular still were only of theoretical interest, however, because of their low melting points and easy solubility.

One day Hill made a crucial observation which dramatically changed the focus of the research. He found he could obtain filaments by using a rod to pull threads from the molten polymer prepared in the still. Furthermore, after these filaments had cooled, they could be drawn to several times their original length. These cold-drawn filaments had quite different physical properties from both the polymer and the "undrawn" filaments pulled by minimal tension from the molten polymer.

The laboratory buzzed with excitement and activity as the group explored the properties of these cold-drawn fibers. X-ray diffraction patterns showed that the filaments in the undrawn state were partly crystalline and that the crystals had a random orientation. The cold-drawn filaments, however, indicated a considerable degree of orientation parallel to the fiber axis. This pattern was similar to that shown in natural silk fibers or in rayon filaments which are formed under tension. The new filaments had high tensile strength and elasticity. They were pliable and tough enough to be tied into hard knots. Drawing the fibers also caused them to develop a high degree of luster and transparency. Moreover, unlike ordinary textile fibers, their wet tensile strength was no less than their dry tensile strength.

Until this time Carothers' research had been entirely fundamental, conceived to explore certain aspects of the polymerization process. However, the striking qualities of the cold-drawn fibers obtained from the polymers aroused hope that it might be possible to make a fiber with commercial utility.

A new research phase began at the Du Pont laboratory—this time to synthesize a polymer that might form the basis for a new marketable fiber. In his earlier research, Carothers had not been successful in



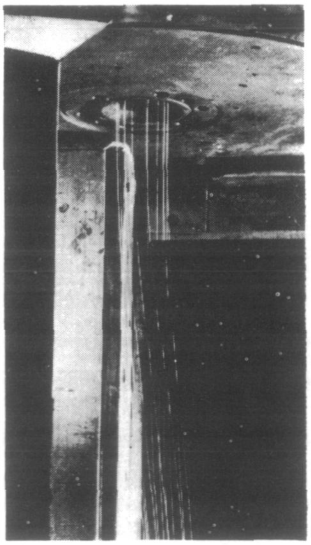
A synthetic fiber was first drawn from a test tube in 1930 by Julian Hill, a member of the research team at Du Pont. (Hill reenacted that historic moment for photographers in 1946.) Five years of intense research was required to produce a marketable product—nylon 66.

Eleutherian Mills-Hagley Foundation, Inc.

preparing a superpolymer with the combination of necessary commercial properties. Surveying his past efforts, he decided to resume work on certain superpolyamides (long-chain chemicals in which the linking group contains carbon, oxygen, and nitrogen). After developing several substances that lacked the desired qualities, Carothers chose to concentrate on a particular polyamide which, when treated in the molecular still, had a melting point of 150° Celsius. After cold-drawing, these fibers were equal to silk in strength and pliability.

After this discovery, Carothers and his colleagues prepared polyamides from a variety of substances and, on February 28, 1935, the first superpolymer from hexamethylenediamine and adipic acid was synthesized. The resulting polymer, poly(hexamethylene adipamide), was designated "66" and later called nylon 66. The first digit indicated the number of carbon atoms in the diamine, and the second digit, the number of carbon atoms in the acid. The new polymer 66 produced fibers that were insoluble in common solvents and melted at about 260° Celsius, providing a margin of safety above usual ironing temperatures. Du Pont selected this particular polymer for development and marketing because it had the best balance of desirable properties and had a low manufacturing cost.

The third phase of research in the laboratory at Du Pont was focused on developing a process for the mass production of nylon 66. To implement this process, it was also necessary to develop the chemical and engineering know-how for the erection and operation of a large scale plant. These tasks were enormous and required the time, talent, and ingenuity of some of the best chemists and engineers of the time.



The spinneret was not a new machine when it was used in the production of nylon. It did, however, have to be redesigned to accomodate some of the unique requirements of nylon production.

Eleutherian Mills-Hagley Foundation, Inc.

Processes for the manufacture of adipic acid and hexamethylenediamine had to be worked out. The machinery and procedures for cold-drawing, sizing, twisting, and packaging yarn had to be planned. In the process of working out all of these elements of the operation, many difficulties were encountered. No previous experience with these types of processes existed in the Du Pont Company, or anywhere else for that matter. Nylon polymer was a completely new material with properties different from those of any other previous synthetic product. Spinning nylon from molten polymer was completely different from spinning either acetate or viscose rayon.

One example of a complex problem which needed to be solved was the development of the pumps to handle the viscous molten polymer. Special pumps had to be constructed which performed under severe temperature conditions, had only small clearances, and used only the polymer itself as the lubricant. A special abrasion-resistant steel which did not warp or soften under these conditions was necessary. In fact, the entire spinning assembly involved radically new engineering accomplishments to produce fibers of the required uniformity and other desirable qualities.

American consumers were, at first, unaware that they were using nylon when, in 1938, toothbrush manufacturers began substituting it for Chinese hog bristles. But Americans awoke to the promise of synthetic fibers at the Golden Gate International Exposition in San Francisco in 1939 when nylon stockings were shown publicly. Buying was frantic when the stockings first went on sale nation-wide about 1 year later. During the first year, retailers sold more than 60 million pairs.

The major advances which allowed the production of nylon did not benefit the American retail market for very long; when the United States entered World War II all American nylon was allocated for the production of war materials. Japanese silk was replaced by American nylon for manufacturing parachutes and government currency. American clothes made from nylon soon disappeared from the market because the synthetic fiber was needed for tires, tents, ropes, ponchos, combat clothes, cargo rigging, and other military supplies.

Although cotton still made up 75 percent of the national fiber market by the end of World War II, there were indications of change. The American public was buying all of the nylon stockings it could, and nylon was being introduced into carpeting and automobile upholstery.

Nylon stockings, first demonstrated early in 1939, were made using experimental production techniques. The full-scale production of nylon was not begun by Du Pont until December, 1939.

Eleutherian Mills-Hagley Foundation, Inc.

Each of these 1,1000 bobbins represent a different item in the nylon product field.

Eleutherian Mills-Hagley Foundation, Inc.

Other Synthetic Fibers

Not every fiber technology currently in American usage came from the efforts at Du Pont. Much of the fiber technology had its origins in other research bases established prior to World War II. In 1893 Edward D. Libbey displayed a fabric containing glass filaments. Scientists from Owens Illinois Glass Company and Corning Glass Works combined efforts to develop a method of spinning relatively fine glass filaments, which eventually resulted in a product that was excellent for insulation and for reinforcing plastics. In 1938 these two companies formed Owens-Corning Fiberglas Corporation. In the early 1960s Owens-Corning produced a very fine glass fiber that was pliable enough to be used in fabrics.

The American Viscose Corporation, this nation's first manufacturer of rayon, was also the first company to spin vinyon into textile filaments in 1939. This fiber, produced today by Avtex Fibers, Incorporated, and the Union Carbide Corporation, is remarkably inert to high concentrations of alkalis and mineral acids. However, it is hard to color and has a relatively low melting point.

Since 1946 when they were first introduced by Doebeckmun Company, metallic fibers (those composed of metal, plastic-coated metal, metal-coated plastic, or having a metallic core), have been used as decorative devices for clothes and home furnishings. Metallic fibers such as Lurex® are found in items such as braids, military uniform decorations, draperies, laces, ribbons, table linens, upholstery, and in even carpets (to reduce the build-up of static charges). Suitable adhesives make it possible to add color to the fibers; and, when properly treated, the fibers are not affected by salt, chlorinated water, or electrolytic or climatic conditions.

In 1948 Union Carbide introduced modacrylic fibers. These fibers have been used to simulate fur and are also found in awnings, blankets, carpets, curtains, scatter rugs, filters, paint rollers, and stuffed toys. Modacrylics remain popular, primarily because they do a good job of emulating the soft, yet resilient, easy-to-dye properties of wool. They are also flame-resistant, quick-drying, and keep their basic shapes under most conditions.

In the early 1950s Du Pont and a few other companies began producing acrylic fibers. Acrylic production represented the successful development of a synthetic fiber which simulates certain characteristics of wool. Acrylics can be used to make a variety of products including sportswear, infant wear, sweaters, work clothes, blankets, carpets, and yarns for hand knitting.

Acrylics and other synthetic fibers that are flexible enough to be joined with other materials made blending with cotton or wool possible. The first result of this mixing was the "wash-and-wear" apparel that appeared in 1952. The original blend was a ratio of 60 percent acrylic to 40 percent cotton. The aim of the blending process was to couple the best characteristics of cotton with those of the synthetic fibers—particularly since it seemed to many experts as though some of the basic advantages of cotton, such as moisture absorption, could not be duplicated.

Polyester fibers, which now represent the single most important class of synthetic fibers, also originated from Carothers' pioneering research at Du Pont. Carothers discovered their fiber-forming properties but, because he was working with a particular type of polyester which he found to be too low-melting and too solvent-sensitive, he redirected his research toward the polyamides (nylon). Subsequently, in 1941, J. R. Whinfield and J. T. Dickson, working for the Calico Printers Association in Great Britain, found that polyesters based on terephthalic acid, an "aromatic" acid, produce very desirable fibers. Their basic patent was acquired by the British company, Imperial Chemical Industries, for all countries but the United States. In a strange twist of fate, Du Pont acquired the United States' right to the commercial fiber which was based on fundamental research done in its own laboratories.

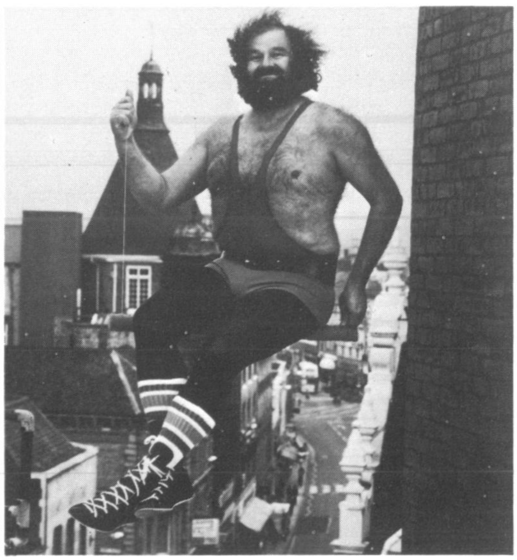
Polyester fibers became commercially available in the United States in 1953. The rapid growth in their production was aided in the 1960s by false twist texturing, a process in which fibers are crimped (or convoluted) to decrease contact with the skin and increase air trapping, thereby increasing comfort. False twist texturing made the popular shape-retaining polyester knits (typically 65 percent polyester and 35 percent cotton) possible and, in addition, remarkably increased the use of synthetic fibers in tufted carpets.

Another significant use for polyester is as fiberfill for pillows and cold-weather clothing. Polyester fibers have largely replaced nylon in tire cords because tires containing polyester not only are strong but also are not subject to the annoying "flat-spotting" that is characteristic of nylon-reinforced tires. Intensive research has produced a gamut of polyester fibers with a broad range of properties, including a type of polyester fiber that can be slowly degraded by the human body and that may, therefore, be used for sutures that need not be removed.

In 1959 Du Pont marketed spandex, an example of a successful "engineered" fiber. It was designed to meet the demand for a stretchable yarn to be used in items such as swimsuits. A stretchable yarn was needed that would be superior to those then available, clear or white in color, easily dyeable, resistant to fading, and stable when exposed to the chemicals added to swimming pools. Later, scientists at U. S. Rubber Company (now Uniroyal, Inc.) developed a spandex in which polyester and urethane were joined.

Du Pont also introduced a unique polymer (an aramid) called Kevlar® which has a combination of high strength and toughness never before produced or found in nature. It is used in bulletproof vests, military helmets, and similar protective clothing. The extraordinary strength of the aramid fibers could not have been achieved without a detailed study of the behavior of the aromatic polyamides when they are dissolved to form solutions. The basic polymers for these fibers are different from the nylon polymers because their chemical constitution makes the giant molecules extremely rigid. A great deal of research was required to find solvents that are suitable for spinning fibers; it is the behavior of the aramid polymers in the solvents that is remarkable. As more and more of the polymer is dissolved, the molecules, because of their rodlike structure, cannot be accommodated in the liquid medium in the typical, random fashion. Instead, when a critical concentration is reached, the molecules rearrange themselves into a crystallike order (a property normally observed only in solids) so that they are more or less parallel to each other. It is this liquid crystalline solution that is spun to form aramid fibers. The molecules emerge from the spinnerette well-aligned with each other and are responsible for the enormous tensile strengths in the resulting fibers. The liquid crystalline behavior of rigid polymer molecules is still an incompletely understood phenomenon, and there are a number of studies underway to correlate the molecular structure of rigid polymers with their properties.

Another Du Pont product is a remarkable polymeric material, Nomex®, which has low flammability and can be used at higher temperatures than most polymers. It is used for ironing board covers, aircraft upholstery, the outfits of professional race-car drivers, and flight suits for aircraft pilots.



The properties of Kevlar® make it an ideal material for use in stronger, lighter bullet-proof vests. A .357 magnum bullet was repelled by the first layer of this Kevlar® vest (left).

To demonstrate the amazing strength of their product, Du Pont had a 280 pound wrestler sit on a trapeze suspended above the main street of Leicester, England by two strands of Kevlar®. Each strand was less than 1/8th of an inch in diameter.

E.I. du Pont de Nemours & Company, Incorporated

Hercules Powder Co. (now Hercules, Inc.) began producing textile-grade polyolefins commercially in 1961, although they had been used earlier for specialized purposes. Polyolefins, derived from propylene and ethylene gases, are characterized by a resistance to moisture and by chemical inertness. Propylene-derived fiber is used for general textile applications.

Optical Fibers

The 1960s were a period in which production facilities of the synthetic fiber industry expanded and efforts that began a full decade earlier to "engineer" fibers continued with renewed dedication. Optical fiber technology became important in the 1970s. A number of techniques developed for the manufacture of synthetic fibers were applicable and helped speed the production of optical fibers. Many key studies that have improved optical fibers related to analyzing the transmission properties of glass, devising better glass formulations, and perfecting commercial scale operations. Mechanical, electrical, and chemical engineers as well as physicists, chemists, mathematicians, and other scientists have all made contributions to this technology.

Fiber optics is that field of physics which deals with the transfer of light from one place to another through long, thin, flexible fibers of glass or plastic; these are called optical fibers. Optical fibers have the ability to transfer light around corners because the sides of the fiber reflect light and keep it inside as the fiber bends and turns.

In 1870 a British physicist, John Tyndall, discovered that light could be guided through a clear substance. At a meeting of the Royal Society for Improving Natural Knowledge, he set up a container full of water with a hole punched in the side of the vessel to allow a stream of water to flow out. Tyndall shone a light down into the container and demonstrated that, as the light came out of the hole, it was guided along the stream of water and lit up the spot on which the water fell. However, it was not until the 1950s that people began to use thin strands of glass to transmit both light and images over short distances.

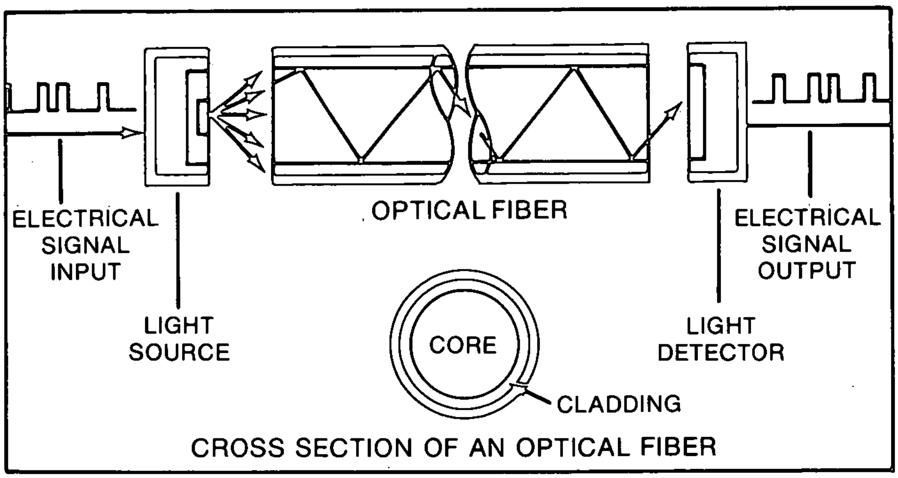
The glass fibers available in 1960 could transmit signals clearly for only about 100 feet. For practical applications, fibers were needed which could carry the light signal more than a mile before needing an amplifier. Obviously, the lower the signal loss, the greater the distance could be between the amplifiers, resulting in increased economy of the system.

The ability of the fibers to transmit light was enhanced by adding a coating of special glass called cladding which kept most of the light waves inside by reflecting them toward the center of the fiber as they passed through it. Researchers also found that the fidelity of the signal could be improved if the fiber were constructed so that its index of refraction (capability to bend light) changed from its center to its outer edge.

The composition of the fiber itself also presented problems. Materials research showed that impurities in the glass produced transmission losses. Metallic ions such as iron were particularly troublesome in this regard, and during the 1970s highly refined processes were developed to eliminate such impurities. Although improvements continue to be made, the losses in signal are now low enough and the quality of transmission good enough that optical fibers can be used for long distance transmission. Bell Laboratories, Western Electric, Corning Glass Works, and International Telephone and Telegraph have all been involved in making these technological advances.

Optical fibers have many uses. They have replaced small electric lights in some medical instruments. Physicians use a bundle of fibers fitted with a lens and an eyepiece to view internal areas of the body. Many dentists have drills which are fitted with optical fibers to concentrate light on the tooth being drilled. In industry, optical scanners are built into data processing equipment to detect the rectangular holes punched in computer cards.

Optical fibers have been adopted by novelty makers for decorations. Bundles of fibers are mounted in a plastic or metal base with white or colored lights shining in the base. The light travels through the glass strands and appears as bright points at the tips of the fibers, thus forming designs.

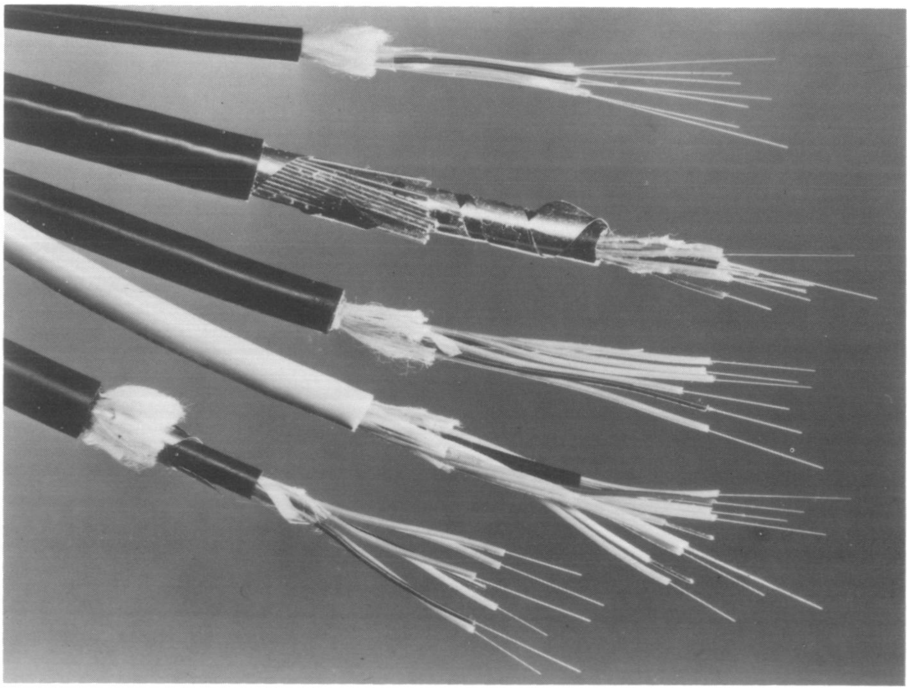


In an optical fiber, electrical signals are converted into light which can be transmitted by the optical fiber; the light can then be converted back into an electrical signal. Cladding helps keep most of the light inside of the fiber by reflecting light waves toward its center as they pass through it.

Currently, however, the greatest potential applications of fiber optics are in communications. It is believed that optical fibers can provide solutions to many of the problems plaguing conventional electronic telephone transmission. Glass fibers' advantage over copper wires lies in the differing natures of electronic and optical communication systems. Electric current used for conventional transmission also creates magnetic fields which distort signals. Other disadvantages of copper include electromagnetic interference, crosstalk, and the need to be amplified at 1-mile intervals by "repeaters."

Because of technological improvements, optical fibers have very recently become competitive with conventional copper cable for some purposes. American Telephone and Telegraph Corporation, General Telephone and Electronics, and other telephone companies now find that it is economically feasible to install glass fibers in their systems, particularly to connect switching centers in urban areas.

A major advantage to these companies is that the optical fiber saves space. A thin fiber cable can carry many more messages than a thick copper cable can. Therefore, telephone companies can add message-carrying capacity to existing underground ducts without having to dig up streets to put in new conduits. In fact, an optical cable with a diameter of 13 millimeters contains 144 fibers that can carry 43,680 telephone calls simultaneously. To carry almost the same



Examples of optical fiber cables for various uses.

Siecor Optical Cables, Inc.

number of calls electronically requires 4 conventional cables comprised of a total of 7,200 copper wires.

Optical fibers have other uses. Electric utilities are starting to employ them for communication and to control circuits within and between generating plants. Because transmissions using optical fibers are relatively immune from "static" associated with high power electrical transmissions, optical fibers are particularly useful near power lines. Before long, optical fiber devices are expected to appear in United States-made cars where they will carry control signals to auxiliary equipment. Cable television companies are using glass fibers on some main circuits. A two-way television system which presently is connected by coaxial cable is being tested in Ohio. The system is called QUBE; it allows people to respond to questions asked over television and to express their opinions instantly from their homes. The high capacity of fiber optics could make it feasible for such a system to be in every home. (A few potential uses of this system are briefly described in the chapter, "Survey Research and Opinion Polls.")

More and more, optical fibers will make their way into wiring systems of aircraft because of their light weight; the armed forces are enthusiastic about this exotic material for many uses. For example, the United States Navy is investigating the use of fiber optics for communicating with submerged submarines. Optical fibers, linked with a radio buoy, would allow messages to be sent without the submarines having to surface. The armed services are also interested in optical fibers for telephone lines because a direct splice must be made to tap light signals, producing a signal leak that can be measured and detected. Copper wire, on the other hand, can be tapped by withdrawing signals from the electromagnetic field that surrounds the wires; optical fibers produce no external electromagnetic fields.

Although the total sales of optical fiber cable and associated equipment currently are relatively miniscule (an estimated $40 million a year), the principal manufacturers expect business to grow to several hundred million dollars a year by 1985.

THE PROMISE OF SYNTHETIC FIBERS

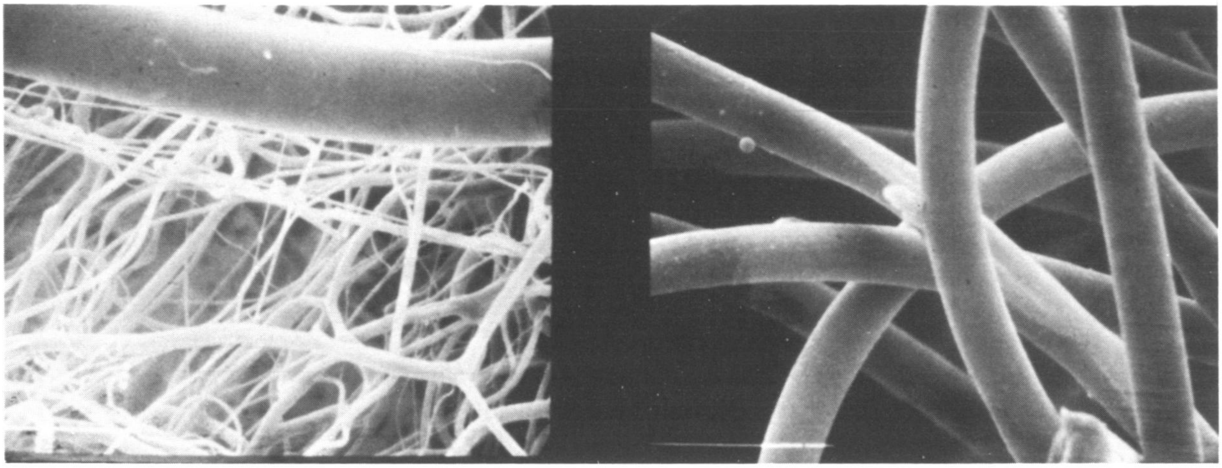
There is much that remains to be done if synthetic fibers are to fulfill society's future needs. Although petroleum prices have skyrocketed, petrochemicals are still the most economical starting material for making synthetic fibers. Approximately 1 percent of the United

States' oil and natural gas production is used to make synthetic fibers. (All synthetic materials consume around 3 percent of the production.) About one-half of that amount is used as raw material for the production of the fibers; the other half is used for energy for the conversion process.

In a world that is dependent upon finite energy resources, how will the fiber industry prepare for the future? Petroleum will never again be regarded as the inexpensive energy source that it once was. Rising worldwide oil prices have influenced the direction of synthetic fiber research away from that of modifying established fibers. Now researchers aggressively pursue ways to replace petroleum as a raw material in the production of synthetics with substances such as coal and biomass (renewable organic resources, such as seaweed, terrestrial plants, garbage, and human and animal wastes). Although the impact of the competition for oil and gas on the synthetics industries is not clear, few experts predict any inexpensive alternatives to these substances as the source for feedstock for synthetic fibers any time soon.

Because of the very rapid rise in the prices of petrochemical feedstock in recent years, some people have predicted a major reversion from synthetic fibers to natural ones. However, there are few signs of this happening. One reason is that many natural materials have also been affected by energy price increases. Another is that synthetic fibers have such widespread acceptance and use in industrial nations. Natural fibers, as currently produced, often require a large input of energy, much of it derived from the same fossil fuels needed to provide feedstock for the synthetics. Cotton is one of the most energy-intensive crops grown. Large quantities of petrochemically based pesticides and fertilizers are needed to grow cotton, and therefore the cost of this fiber has been directly affected by price hikes for diesel fuel. The spinning of cotton yarn and weaving of cotton fabrics are highly mechanized in developed countries. In a large part, the increased productivity in cotton mills over the last three decades has been made possible by the substitution of cheap energy for laborers who used to do much of the work by hand.

The production of textile fibers provides an interesting illustration of the comparative energy requirements of synthetic and natural materials. Cotton and polyester frequently may be substituted for one another, and their relative use is often determined by price. One study completed in 1978 concluded that 25 percent more energy was required to make a polyester/cotton blend shirt than to make one entirely of cotton. However, cotton/polyester fabrics are more durable and require less maintenance than all-cotton fabrics. The life cycle energy



This comparison of 3M's Thinsulate® (left) and polyester (right) shows how Thinsulate's® micro-fibers can trap more air in a given space, resulting in a more efficient insulator.
3M

requirements of all-cotton clothing in the United States are estimated to be as much as 90 percent higher than for the blends—though this varies considerably depending on individual practices of washing and ironing clothes. In industrial countries where many people have electrical washers, dryers, and irons, it is likely that polyester/cotton blends actually save energy. In developing countries the opposite may be true.

One factor often overlooked is how closely linked the international food and fiber markets are. The world food crisis in the early 1970s caused many farmers to switch from fiber to food crops, and the generally tighter market increased cotton prices by more than 50 percent. This increase seriously damaged cotton's competitiveness with synthetics. The production of cotton, which had been growing slowly, leveled off. Cotton producers in America were able to adjust quickly by shifting from growing cotton to growing soy beans and other crops, but few farmers in developing countries had that latitude.

Thinsulate® provides approximately twice as much thermal resistance as down does; therefore, Thinsulate® is often used in winter outerwear.
3M

There are ways in which synthetic fibers themselves are used to reduce the consumption of energy. Technologists have developed a new material composed of a combination of polyester and extremely fine olefin fibers for use in insulation. It has just recently entered the commercial market. Thinsulate® produced by the Minnesota Mining and Manufacturing (3M) Corporation, is twice as warm as wool, with only a fraction of the thickness and bulk. The scientific principle behind Thinsulate® is that micro-fibers trap more air in a given space than do fibers of a larger caliber. Glass fiber insulation in buildings is a well-known aid to energy conservation.

Research into the properties of some fibrous materials may result in the development of less expensive, more advanced, and more effective devices to filter out polluting particles. These filters will permit other reserves, such as coal, to be used more readily, with less fear of environmental contamination.

Many of the research and development activities concerning fiber-reinforced composites also are directed toward finding ways to save energy. Generally speaking, these substances are masses of fibers dispersed in a polymer or metal matrix, and they are becoming a new class of engineering materials. These products attempt to reproduce the intricate structural networks found in nature which provide so many astounding functional characteristics.

The automobile industry is trying to produce a new generation of vehicles that have significantly improved fuel economy. One key to achieving this goal is to develop strong, economical, and light-weight materials to use in automobiles. It may be possible to reduce the weight of a vehicle more than 30 percent by using parts made of fiber-reinforced polymers.

The reduced weight advantages of composites are particularly important in high-performance aircraft and in military and space vehicles. Commercial aircraft now are being refitted with these lighter materials to enhance fuel efficiency. For example, the next generation of Boeing aircraft (the 757 and the 767) will make significant use of graphite fiber and graphite/Kevlar® fiber reinforced composites.

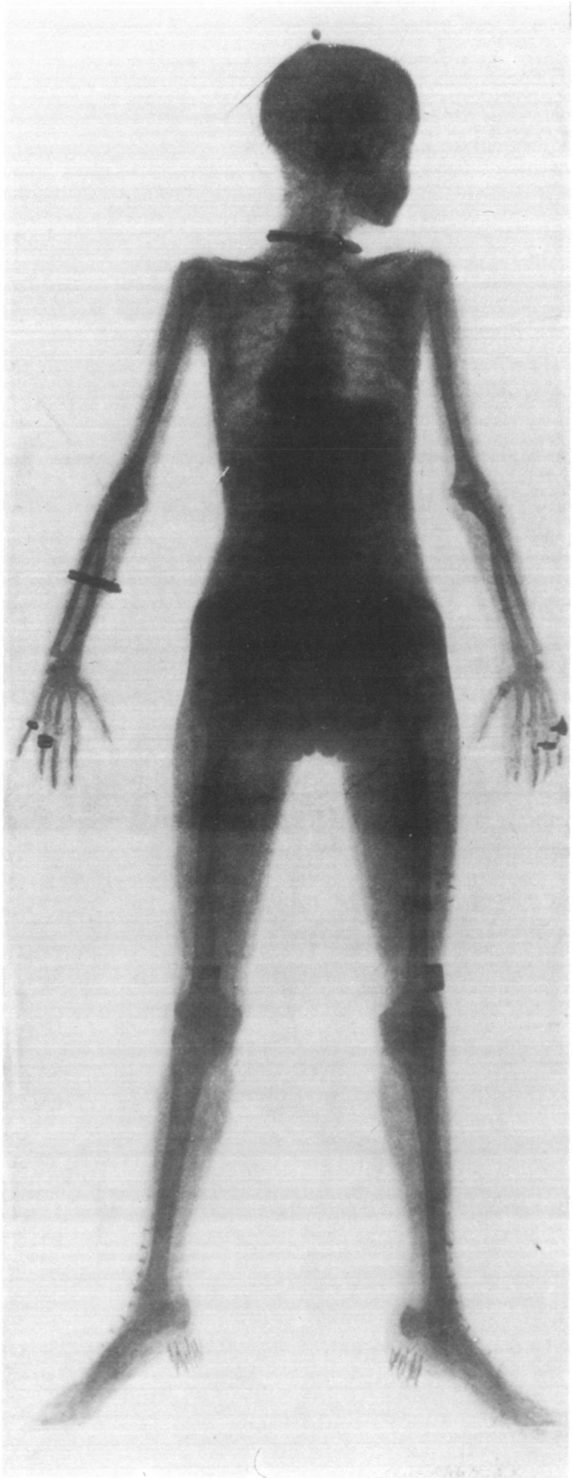
In addition to research prompted by energy shortages and the search for synthetics with the desirable properties of wool, cotton, and silk, there are other driving forces for more study in the field of synthetic fibers. Static electricity in clothing and carpeting made from synthetics is still a problem. Methods of synthetic fiber production can be refined to give even greater savings to the consumer. Dyeing and finishing processes as well as soil resistant and repellancy qualities can be improved. The expanded use of glass fibers in telecommunications will require investigation and technological innovations.

There are new methods which make textiles resist burning. More and stronger regulations will undoubtedly increase the consumption of flame retardant materials and will encourage studies on how to reduce flammability of synthetic fibers.

The synthetic fibers industry has been firmly founded on basic research discoveries. The science of giant molecules, or polymers, is a relatively young field and is expanding in the United States, Japan, West Germany, Great Britain, France, and the Soviet Union. Although many of the key discoveries were made by industrial scientists, the base of research in polymer science in the United States is shifting toward universities. Several large coordinated programs of polymer research have been established at the University of Massachusetts, Case Western Reserve University, and the University of Akron. Other universities such as the University of Wisconsin, the University of Michigan, Massachusetts Institute of Technology, the University of Utah, and the Polytechnic Institute of New York are major centers of polymer science and engineering.

Currently, the synthetic fiber industry enjoys a promising future, witnessed by the strength of the American companies mentioned in this chapter. All are publicly held and traded corporations, and the rate and volume of investments in these firms provide some indication of the general health and public confidence enjoyed by the industry. In 1976 investments in synthetic fiber plants in this country totaled $8.2 billion. Synthetic fiber sales jumped from $5.4 billion in 1977 to $7.4 billion in 1979. In 1975, some 100,000 employees, at an average salary of better than $16,000 a year, were making and marketing synthetic fibers.

The United States still leads the world in both synthetic fiber production and consumption, but Japan and some Western European countries have patterns of synthetic fiber research, development, and increases in production that resemble those followed by the United States in years past. Synthetic fibers now represent about one-third of the world fiber market and their consumption is increasing annually at the rate of about 9 percent worldwide. There is little doubt that the world market already reflects the many advantages of synthetic fibers. And given the challenges of the future, synthetic fiber technology is bound to increase.



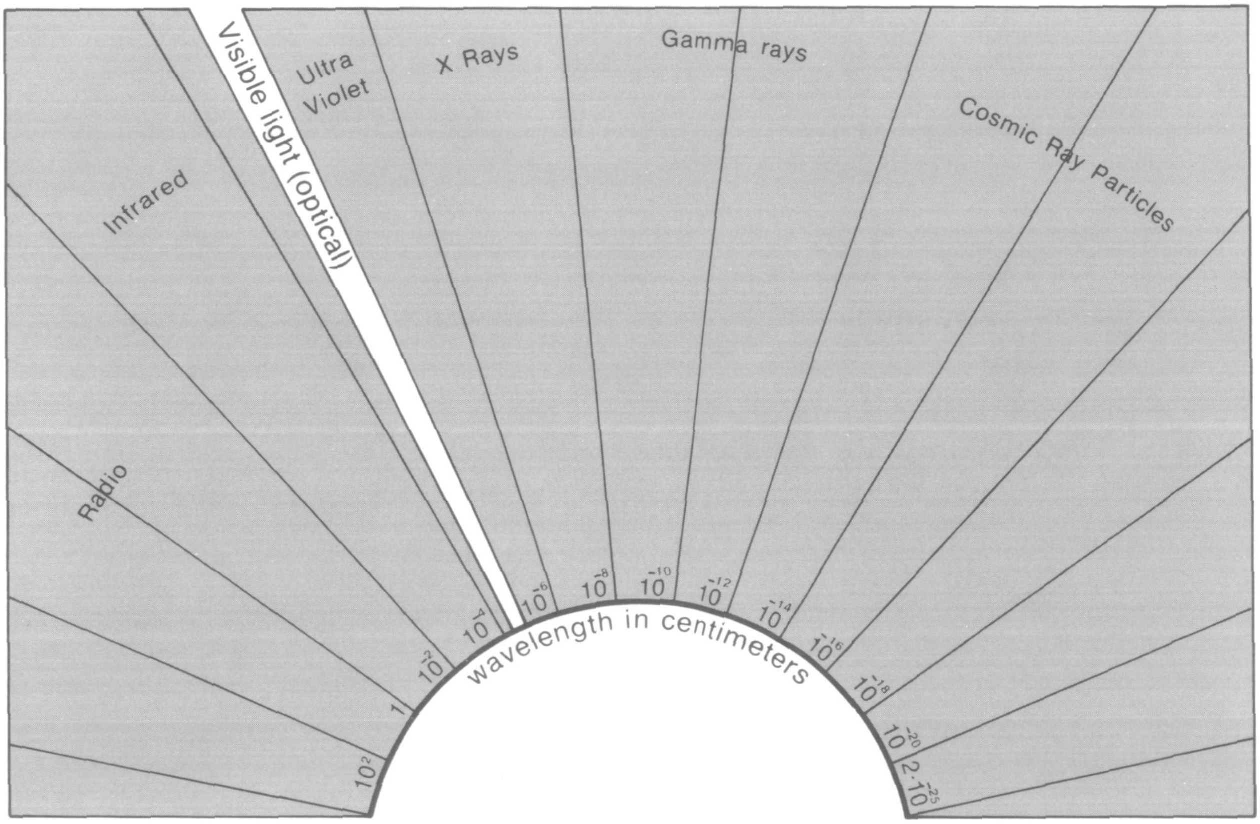
This is the first radiograph of an entire adult body taken at a single exposure. The subject was exposed to X rays for 30 minutes. Her pin, necklace, bracelet, rings, and high-button boots with nailed-on heels are clearly visible. The radiograph was taken in 1897 by W. J. Morton, the son of William Morton, who in 1846 first used ether for surgical anesthesia.

National Library of Medicine

X rays for Medical Diagnosis

Many people think of radiation as a relatively new phenomenon—a product of the nuclear age—but radiation has been around as long as the universe. In fact, sight is possible only because of light rays, which are actually "visible" radiation; heat from the sun is another type of radiation. Everyone is familiar with a rainbow—the beautiful array of color resulting when sunlight is split as it passes through droplets of water. Just as the colors of the rainbow are constituent colors of visible light, so a number of other radiations are constituents of the much broader range or spectrum of so-called electromagnetic radiations shown on page 180. Radio waves, which are at one end of the spectrum, have very different characteristics from X rays, which are at the other. But electromagnetic waves have common properties. They are all rapidly fluctuating electric and magnetic fields that travel through space with the same velocity—about 186,000 miles per second, the speed of light.

Electromagnetic waves differ in how they vibrate as they move away from their sources, and they also differ in energy. High-energy waves vibrate more frequently than low-energy waves. The wavelength, or distance between corresponding parts, of high-energy waves is shorter than that of low-energy waves. In other words, the higher the energy, the shorter the wavelength. Electromagnetic waves carry momentum and can transfer momentum the same way particles do. In many circumstances, radiation can be thought of as a stream of particles, called photons, traveling at the speed of light and exhibiting wave-like phenomena. For comparison, visible light photons have energies of from about 2 to 3 electron volts, but X-ray photons have energies ranging from 125 to 1,250,000 electron volts. The photons at this high-energy end of the electromagnetic spectrum are called X rays



This diagram of the electromagnetic spectrum illustrates the difference in wave lengths between visible and non-visible rays, including X rays.

and gamma rays. The only difference between X rays and gamma rays is in the way in which they are formed. In most instances, gamma rays are emitted by spontaneous intranuclear decay of radioactive isotopes, and X rays are produced by transitions in the inner electron energy levels of an atom and are usually induced by electron bombardment of a metallic target in an X-ray tube. Once a photon is generated in either manner there is nothing to distinguish its origin.

The practice of medicine has been greatly influenced by the application of X rays to the diagnosis of illnesses and injuries and to the treatment of certain diseases. Approximately two-thirds of the population of the United States is exposed to X rays for medical or dental purposes each year. This chapter will concentrate on the uses of X rays in medical diagnosis. It will describe the discovery of X rays; the development of some of their diagnostic uses; several experiences of scientists and clinicians with X rays; and a few milestones in the evolution of the technology of diagnostic X-ray equipment. Some historical material in this chapter is based on descriptions in *The Rays*, a book prepared by Ruth and Edward Brecher, at the invitation of the American College of Radiology Foundation.

The use of X rays in medical diagnosis is a broad subject. It is, therefore, beyond the scope of this chapter to include discussions about radioisotopes which emit gamma rays; the therapeutic uses of radiation; or other important methods of diagnostic imaging, such as ultrasonography, thermography, nuclear magnetic resonance (NMR), dynamic spatial reconstruction (DSR), or positron emission transaxial tomography (PETT). Although the long-term and short-term hazards of excessive radiation are also very important subjects, they will not be a primary focus either. Hazardous effects are well-recognized, and procedural and technological progress has been made to minimize the dangers to millions of patients, to those who use X rays in their work, and to the public. For example, a patient getting a routine diagnostic chest X-ray photograph in the early 1940s would have received a radiation dose more than 10 times that to be expected today. The reductions of dosages used in fluoroscopy have been even more impressive; in the same time period, new procedures have made it possible to reduce radiation exposure during fluoroscopy to as much as 1/100 of the exposure that was common in the early 1940s. Dosages to the ovaries or testes, when other parts of the body are being X-rayed, have been reduced more than 400 times. All of these reductions in dosage were achieved during the time in which technological improvements accompanied improvements in the quality of the X-ray films. Scientists, clinicians, and the general public all recognize that it is important to keep human exposure to radiation at the lowest possible level; and a great deal of thought, research, and legislation continue to be directed toward that goal.

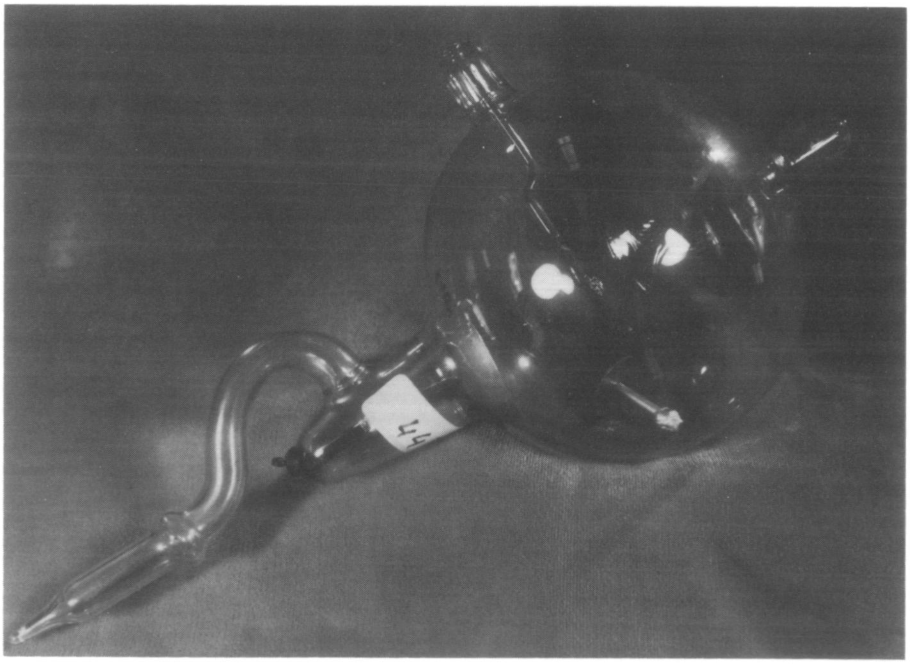
HISTORY

Goodspeed, Jennings, and Crookes

Sir William Crookes
(1832-1919)
American College of Radiology

In 1890, nearly 6 years before X rays were discovered, two men, each 30 years old, met in a physics laboratory in Philadelphia. Arthur Willis Goodspeed was a professor of physics; his visitor was a photographer, William Jennings. Jennings had brought his photography equipment with him to the laboratory. For some time he had wanted to experiment by making photographs using light from an electric spark, and Goodspeed had agreed to help him.

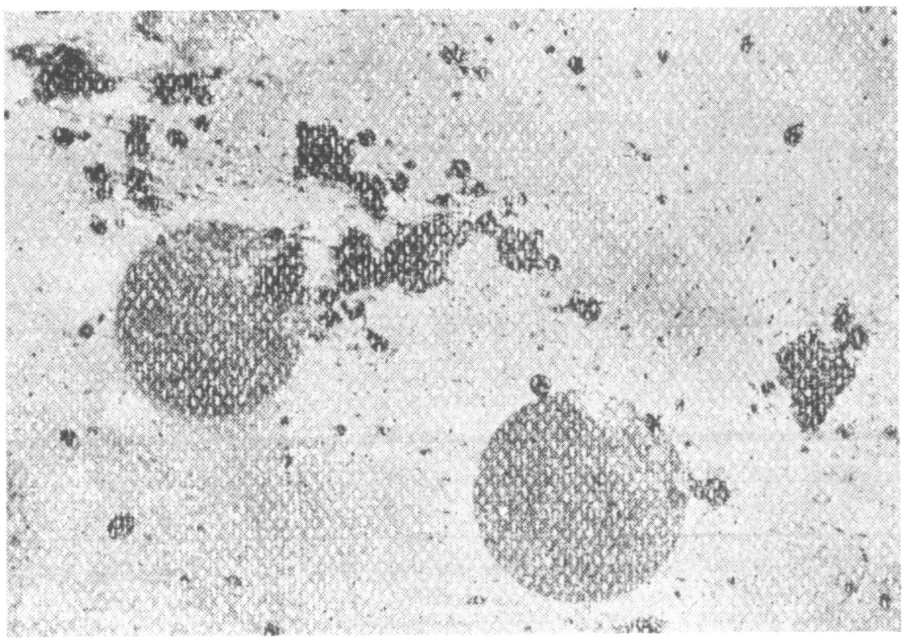
The two men set up an induction coil (to convert low voltage into high voltage) which was connected to a spark gap. Jennings placed his photographic plates so that light from the sparks generated by Goodspeed's equipment would make it possible to photograph coins and other items placed on the plates. Goodspeed's induction coil was the one he ordinarily used to demonstrate how a "Crookes tube" functioned. A Crookes tube, devised in 1875 by Sir William Crookes, was essentially a glass tube from which most of the air had been evacuated. The negative terminal of a high voltage source was connected to an electrode (the cathode), and its positive terminal was attached to another electrode (the anode). When current was applied, a stream of electrons, usually referred to as "cathode rays," flowed from the cathode to the anode and made the glass of the tube fluoresce.



The Crookes tube was devised in 1875 by Sir William Crookes.
American College of Radiology

That evening, after the photography session was over, Goodspeed also demonstrated several of his Crookes tubes to Jennings. Neither man noticed that a stack of unexposed photographic plates, left over from their earlier experiments and still wrapped in light-proof paper, was lying nearby. Two of the coins which they had been photographing lay on top of the stack. When Jennings left the laboratory, he took both the exposed and the unexposed plates home with him.

Several days later, Jennings told Goodspeed that something mysterious had happened. He had noticed "fogging" of some of the unexposed plates. When he had developed one of these plates, it showed the outline of two unexplainable small round objects. Neither man could account for what he saw, and the matter was forgotten for nearly 6 years. Jennings, however, was sufficiently curious about the disc-shaped images to file the plate safely away. In fact, that X-ray plate from February 22, 1890 is the oldest known North American X-ray plate for which a record survives.



Arthur Willis Goodspeed's X-ray photograph made on February 22, 1890.
National Library of Medicine

Crookes himself (and undoubtedly others) had a similar experience. In his address to the Roentgen Society in 1897 Silvanus Thompson reported that, "Crookes, after experimenting with his tube and developing some photographic plates, noted on his pictures marks that corresponded to his fingers; thinking the plates were defective, he returned them to the manufacturer with some strong remarks!"

Lenard and Roentgen

Between 1890 and 1895 many people studied the cathode rays generated in the Crookes tube; however, most of their experiments were frustrating to the investigators because studies of cathode rays (known today as beta rays or electrons) could only be conducted within the cathode-ray tube itself. Then a German scientist, Philipp Lenard, developed a tube which had a window covered by aluminum foil. He found that the cathode rays penetrated the window and that their effect could be observed on photographic plates several inches away. The rays could also be seen faintly emerging through the window if the room were dark and the glass portion of the tube had a lightproof cover on it.

In the 1800s many scientists were interested in studying cathode rays. One of them was Wilhelm Conrad Roentgen, the director of the Physical Institute at the University of Würzburg. On Friday, November 8, 1895, Roentgen was beginning to conduct an experiment similar to the early ones of Goodspeed and the later ones of Lenard. As in Goodspeed's experiments, Roentgen used an ordinary Crookes tube connected to an induction coil. As in Lenard's research, the room was dark. But unlike Lenard's experimental conditions, the Crookes tube had no aluminum window and was completely surrounded with thick cardboard. (Roentgen's earlier experiments with the Crookes tube had been done with the glass uncovered so he could watch what was going on inside the tube.) It so happened that in the darkened laboratory Roentgen's newly acquired barium platinocyanide screen (which could detect the presence of cathode rays outside of the tube) was on a counter several feet away. One of Roentgen's biographers, Otto Glasser, has described what happened when Wilhelm Roentgen activated the lightproof tube in a pitch black room:

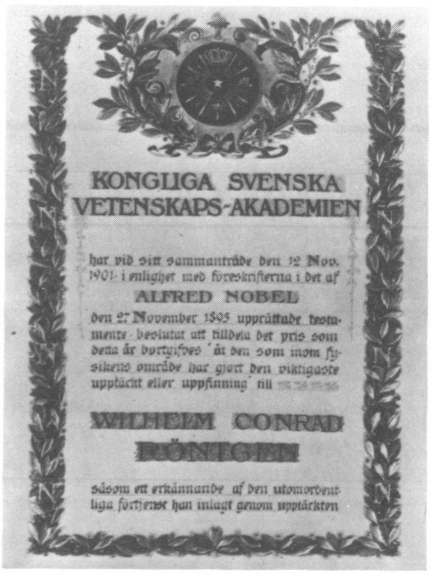
Wilhelm Conrad Roentgen (1845-1923)
American College of Radiology

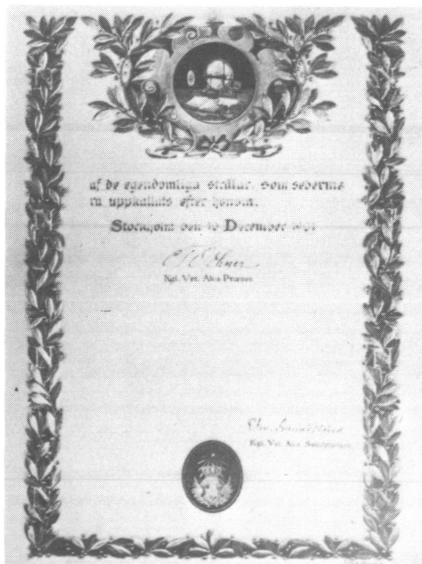
> Suddenly, about a yard from the tube, [Roentgen] saw a weak light that shimmered on a little bench he knew was located nearby. It was as though a ray of light or a faint spark from the induction coil had been reflected by a mirror. Not believing this possible, he passed another series of discharges through the tube, and again the same fluorescence appeared, this time looking like faint green clouds.... Highly excited, Roentgen lit a match and to his great surprise discovered that the source of the mysterious light was the little barium platinocyanide screen lying on the bench. He repeated the experiment again and again, each time moving the little screen farther away from the tube and each time getting the same result.

Roentgen's screen lit up even though the cathode rays could not escape from the tube because it had no aluminum window. If the cathode rays could not have illuminated the screen 6 to 7 feet from the tube, what had caused the screen to glow? Clearly what Roentgen had seen was not the result of visible light or of cathode rays—he had discovered a new kind of radiation. Because at first Roentgen did not understand what the rays were, he called them "X" rays because "X" symbolizes the unknown.

To record what he had seen that afternoon, Roentgen placed a photographic plate in the path of the rays and captured proof of his observation. Subsequently, he put a variety of substances between the tube and the photographic plates. He found that the exposure of the plates changed when he used different materials and that the exposure also varied with the thicknesses of the materials. Roentgen saw that if a hand were held before the fluorescent screen, bones made a dark shadow and the surrounding soft tissues produced only vague outlines.

On December 28, 1895, Roentgen submitted a manuscript describing his findings to the Physical-Medical Society of Würzburg. Soon Roentgen's discovery became front page news around the world. Scientists began clamoring for commercially available Crookes tubes, and when the supply ran out, investigators began to make their own. Within the first year after Roentgen presented his findings, more than 1,000 scientific papers and 50 books were published dealing with X rays. For his pioneering work Roentgen was awarded in 1901 the first Nobel Prize shown below.





The first Nobel Prize awarded to Wilhelm Conrad Roentgen in 1901.

An early X-ray photograph made by Wilhelm Conrad Roentgen in late in 1895. It is speculated that this is the hand of Roent-gen's wife.

American Institute of Radiology

By early January 1896, Roentgen's pioneering paper was in the hands of scientists throughout Europe. They were able to reproduce his results quickly. Late in January investigators in the United States followed suit. Then, at almost identical times, Arthur Wright of Yale University and John Trowbridge of Harvard University reported successful experiments, which were followed soon thereafter by dozens of reports of others. Within a very short time there was widespread use of X rays in medicine. They were used to locate foreign objects, to examine broken bones, and to do procedures that are often regarded as modern developments. In fact, at the 1896 meeting of the Association of American Physicians, reports were presented on the use of X-ray plates to study the blood vessels in an amputated hand, to examine a patient's esophagus following a swallow of nitrate of bismuth (a liquid which can be seen by using X rays), and to locate an aortic aneurysm (a dangerous abnormal ballooning of the largest artery in the body).

The first X-ray photograph of an entire adult body was published in 1897 and is shown on page 178. The subject in the picture was still fully clothed; and her necklace, bracelet, rings, hat pin, and high-button shoes with nails in the heels are clearly visible. In making this particular picture, the subject was exposed to X rays for 30 minutes.

Hazards Recognized

While researchers in Europe and the United States were finding and recounting multiple beneficial uses for X rays, some disconcerting stories began to be circulated. Evidence was accumulating that, as the rays passed through human tissue, they could cause damage. It was not long before reports of serious harm from X rays began to appear.

The *Electrical Review* reported one such case in August of 1896. A recent graduate of Columbia College, H. D. Hawks, had been working for about 4 hours a day around an unusually powerful X-ray machine. After 4 days, he had to leave work because his hands began to swell and to appear very sunburned. Two weeks later, all of the skin came off of his hands, his fingernails stopped growing, and the hair on the sides of his head began to fall out. The hair at his temples disappeared because he routinely placed his head near the tube to demonstrate his jawbone to visitors who were curious about X rays. His eyes were very bloodshot, and his vision was impaired. After being treated by physicians for the severe burns and recovering somewhat, Hawks went back to work on the X-ray equipment. Following the trial of a number of unsuccessful methods, he discovered he could protect his hands from the X rays by covering them with tinfoil.

There were other reports of damage caused by X rays during 1896. One equipment manufacturer, G. A. Frei, proposed that it was not the X rays that were the cause of the problems:

> Many physicians bring forth the argument that the application of the X rays might prove dangerous to their patients, that here a foot had to be amputated, there someone's fingernails dropped off, another had a sore of three months' standing, etc. Such arguments can be met with the above fact that the X-rays are not the direct cause of the trouble, and with this fact established, remedies could undoubtedly be found to reduce, if not entirely eliminate, the effect on the skin when coils are to be employed.

Frei claimed that it was the induction coil rather than the X rays that made the problem. At the time, he was selling static machines to produce the high voltage electric potential necessary to make X rays. These were the same machines used by salespeople to demonstrate the effectiveness of lightning rods.

A physicist at the General Electric Company, Elihu Thomson, had developed an induction coil especially for X-ray use that could be used instead of the static machines made by Frei. Thomson set out to prove that X rays (not the induction coil) caused the damage by deliberately making crucial experiments on himself with the intention of publishing the results. He asked himself what part of his body he could best afford to lose and decided it was the last joint of his left little finger. He repeatedly exposed the fingertip to an X-ray tube excited by a static machine, shielding the rest of his hand with glass. Eleven days after the exposure, the skin on the back of his finger had begun to blister and was "red, swollen, and painful to touch." A month and a half later his finger was still sore. Thomson had proven that the burns were a direct result of the rays themselves and the induction coil was not the cause. A year later, Thomson performed another experiment. He exposed a different finger to X rays for only a short time each day for several days and demonstrated that the harmful effects were cumulative. By 1898 Thomson's reports convinced most people working with X rays to use lead shielding and to wear protective gloves. His work apparently was not convincing enough for at least one of Thomas Edison's employees.

Thomas Edison

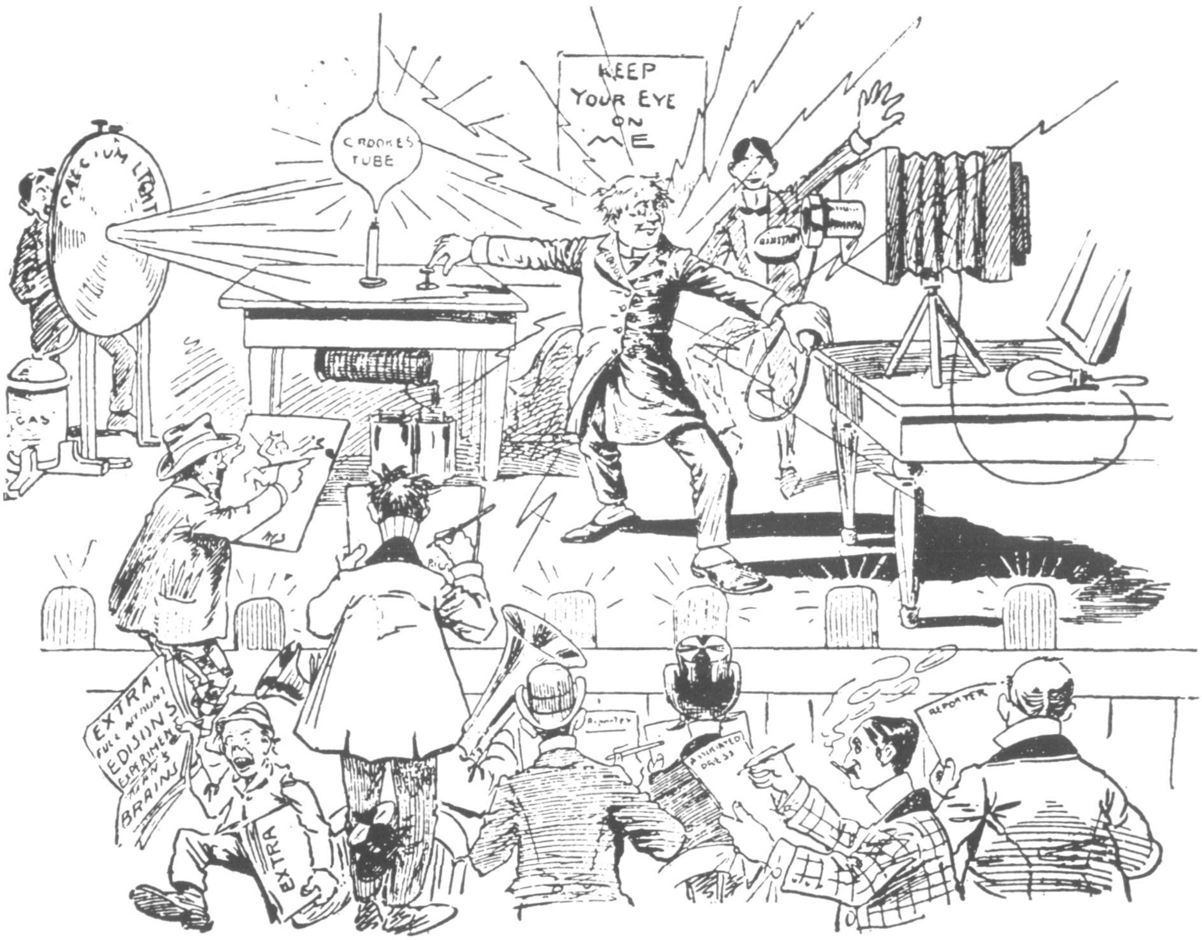
Roentgen's discovery stimulated Thomas Edison's curiosity, and in February of 1896 the New York *World* quoted Edison as saying, "When I have a sufficiently powerful engine I'm sure that there will be no question of obtaining a good photograph of a man's hand .... I expect

A promotional photograph taken during Thomas Edison's attempt to take an X-ray photograph of a human brain.

to have my dynamo in operation by Tuesday (February 4) at the latest." Edison was also fascinated with the idea of his being the first person to "photograph" a living human brain. On February 5, he was challenged by a telegram from William Randolph Hearst, then publisher of the New York *Journal*:

> WILL YOU AS AN ESPECIAL FAVOR TO THE *JOURNAL* UNDERTAKE TO MAKE CATHODOGRAPH OF HUMAN BRAIN KINDLY TELEGRAPH ANSWER AT OUR EXPENSE.

A number of reporters and onlookers were attracted to West Orange, New Jersey, where Edison had announced he was going to photograph the living human brain (Figure 0). An assistant recalled, "For three weeks, more than twenty newspaper reporters were stationed at the Laboratory, the work going on nights, days, and Sundays." The *Electrical Review* said, "He had the 'boys,' as he calls his assistants, working most of the time. When Saturday night came, Edison had been working steadily for the better part of 70 hours, so he went home to rest on Sunday. Bright and early Monday morning he appeared at the laboratory as chipper as a lark." The *Electrical World* reported that "Edison himself has been having a severe attack of the Roentgen mania." It also told of Edison's using the music from a loud hand organ to keep his staff awake during the long hours they were asked to work.



A cartoonist's impression of the scene at Edison's laboratory in East Orange, New Jersey.
from "The Trail of Invisible Light" by E.R.N. Grigg, M.D.

But the scheduled photographing of the brain did not take place on February 8. Edison explained by saying he was working on some special tubes for the experiment. By February 11, Edison still had not taken a photograph of a human brain, but even that delay was regarded as important by the *Times*, which reported that the tubes "were not ready. He spent a busy day, however, in making experiments designed to bring out the properties of X rays."

It was not long before almost everyone in the country was following Edison's progress, and there was daily reporting of the activities at his laboratory. By February 12, reporters were getting restless. During one experiment they watched a photographic plate exposed for a full hour, and when it was developed nothing could be seen but a murkiness and a curved line which Edison could not explain. Edison's apology was quoted in the *New York Times* on February 13: "A man making experiments may count himself lucky if he has successful results in ten out of every hundred experiments which he makes. At the same time, each negative result obtained under proper conditions closes up some avenue along which no further experiments are needed."

There were more delays and more statements. On February 23, Edison was recounting the misfortunes he had had: "My tubes have burst, I have not been able to get as high a vacuum as desired, and I have had to substitute two Leyden jars for the condensers which I have used heretofore.... I tried [a] German glass [tube] ...and at first got good results.... Afterward it failed to act altogether." The show went on, but the audience began to disappear.

So far as is known, Edison never did succeed in producing that promised X-ray plate of a human brain. He did, however, manage to have work of some importance going on in his laboratory while he was out front performing for visitors. He made improvements in the Crookes tube and became known as the inventor of the fluoroscope. In fact, Edison did not actually invent the fluoroscope; he did, however, make several contributions to its evolution. He developed a fluoroscope which worked like an ordinary stereoscope which allowed both eyes to focus on the screen at the same time. He coated the screen of his fluoroscope with calcium tungstate to make the screen much more sensitive than one coated with barium platinocyanide.

Edison's most important contribution to fluoroscopy, however, may have been economic. He turned the manufacture of the fluoroscope over to a company which made it available at a relatively low price; as a result, the fluoroscope became commercially available as early as March 1896.

Not long after Edison's unsuccessful attempt to photograph the human brain, he became interested in fluorescent light bulbs. He began using modified Crookes tubes as light bulbs, calling them "fluorescent lamps." According to the account of E. R. N. Grigg, the results were disastrous: "I [said Edison] started in to make a number of these lamps, but I soon found that the X ray had affected poisonously my assistant, Mr. Dally, so that his hair came out and his flesh commenced to ulcerate. I then concluded it would not do, and that it would not be a very popular kind of light, so I dropped it."

Clarence Dally, who was Edison's glass blower, died in 1904 before reaching his 40th birthday. It is most likely that he was the first fatality from X-ray "poisoning" in the United States.

Evolution of a New Specialty

A logical consequence of the rapidly growing number of uses of X rays in medicine was the emergence of a small group of physicians who were interested in using X rays in their practice. At a meeting of the American Roentgen Ray Society in December 1908, Percy Brown said that, 10 years before, he had heard that "any high school boy, given a smattering knowledge of human anatomy, should be able to 'take X-ray pictures,' to borrow the expression used." He went on to say that in the early days all that was needed to take X rays "was a hand to start the machine and then—to stop it. A process purely mechanical; no brains necessary." However, he pointed out that in 1908 physicians in isolated communities had to struggle to determine the position of fracture fragments with a complete disregard for certain fundamental laws of optics or in blissful ignorance of the hazards of improperly used X rays, and in many cases the patient regarded the cure as worse than the disease. And then there were still the dangers from frequently missed diagnoses.

One physician attending the meeting explained his experience with X rays:

> I was one of those who early tried the roentgen rays. I found, before the lapse of many years, that if I was to do as good X-ray work as was being done by others, I could not practice surgery. I had no time to develop my plates except at night, when I required sleep. The demands of my surgical practice were too exacting to permit me to do justice to the X-ray work.

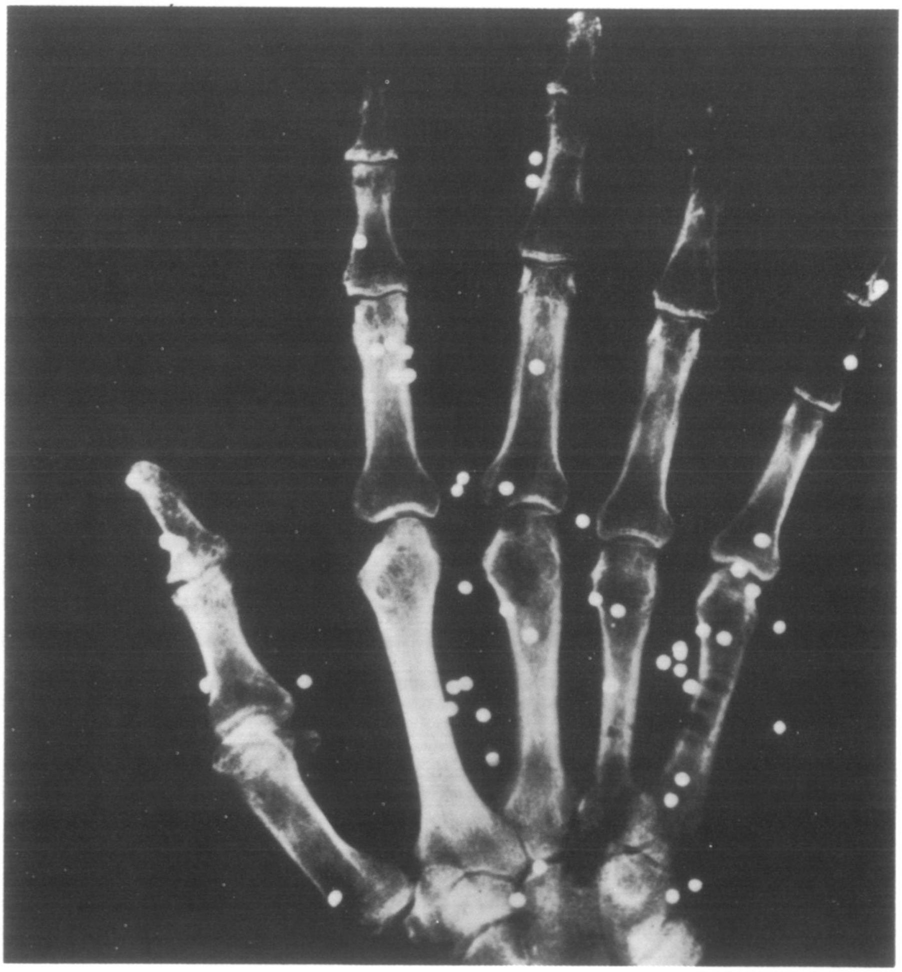
Some physicians in similar situations also chose to refer their work with X rays to other physicians who shifted the focus of their work to become "radiologists" or "roentgenologists." By 1910, 27 percent of the members of the American Roentgen Ray Society who responded to a questionnaire were involved in "practicing roentgenology as an absolute specialty," and 90 percent of the members responding thought that the demand for X-ray work was growing.

A very interesting finding from this same survey was that there was a considerable division of interests already developing among those who practiced radiology. Forty-nine percent of those who responded considered their practice to be primarily diagnostic, and 19 percent categorized their use of X rays as only for therapy. Only 30 percent of the members of that Society responded that they were using the rays for both therapy and diagnosis. (Two percent engaged in neither diagnostic nor therapeutic use of X rays.)

It is noteworthy that as recently as the 1950s and 1960s radiologists and medical educators were arguing whether or not radiology should be divided into two independent specialties—one concerned with radiation used to diagnose medical problems and the other specialty dealing with the use of radiation as a form of treatment or therapy.

DIAGNOSTIC RADIOLOGY

With X rays it was not difficult to localize certain foreign substances in the human body, such as a bullet in the hand or foot. X-ray studies of bone fractures also were common and very useful. But there were radiologists who wanted to extend their investigations into other areas of the body, such as the brain, the gastrointestinal tract, and the cardiovascular system.



This early X-ray photograph shows the distribution of buck shot in a human hand.

from American Institute of Radiology

The Gastrointestinal Tract

Physicians who treated stomach and intestinal disorders at first found no reason to be excited about X rays. The stomach and the small intestine produced almost no detectable X-ray shadows. The researchers were aware that they would have to devise techniques and identify signs or shadows which had diagnostic significance. Many radiologists were using the method which they later called "retrospectroscopy." This method required that the radiologists be present when the patient was examined at autopsy or had surgery, to learn the diagnosis, and that they re-examine the patient's X-ray photographs to look for previously unsuspected diagnostic signs which could also be found in other patients with the same condition.

The key to X rays' usefulness for the gastroenterologist was the discovery that, if a substance opaque to the X rays could be swallowed or introduced into the bowel by enema, the intestinal tract could be "visualized."

Walter B. Cannon (1871-1945)

From Rigler, Leo G., and Weiner, Marvin: History of roentgenology of the gastrointestinal tract. In Margulis, Alexander R., and Burhenne, H. Joachim, editors: Alimentary Tract Roentgenology, ed. 2, St. Louis, 1973, The C. V. Mosby Co.

In fact, as early as 1896 Walter Cannon experimented with a goose and "made for it a box so arranged that the long neck reached up through the cover. A high cardboard collar was then attached to the top of the box in such a way that it could be closed in front and surrounding the goose's neck. Thus the goose, with the appearance of using the most stylish neckwear, presented to the fluorescent screen a very satisfactory extent of esophagus." At a professional meeting that same year, the phenomenon of swallowing was demonstrated informally by having the goose swallow capsules containing bismuth subnitrate. This probably was the first public use of X rays to demonstrate the movements of the digestive tract.

It had been known for some time that extreme emotions could affect the digestive process. When Cannon was still a freshman at Harvard Medical School, he was using X rays to observe the repetitive contractions of the stomach of a cat to which he had fed a meal of bread mixed with bismuth subnitrate. While he was watching, the cat "suddenly changed from her peaceful sleepiness, began to breathe quickly, and struggled to get loose. As soon as the change took place, the movements in the stomach entirely disappeared." He continued his observations and calmed the cat. As the cat relaxed and began to purr, the rhythmic movements in the stomach began once again. Cannon also found that if he made the cat uncomfortable by holding her mouth closed and covering her nostrils to keep her from breathing, the movement of the stomach would stop. When the cat was allowed to breathe normally, the gastric activities reappeared. These studies done long ago are still of value to modern radiologists because these studies help in distinguishing disorders of the stomach which are related to anxiety from those which are organic.

As early as 1897 Charles Lester Leonard reported that he had diagnosed a case of drooping of the stomach by introducing an emulsion of bismuth into a patient's stomach. He said that the stomach had drooped so low that part of it was visible through the pelvic bones. At a medical meeting that same year, he explained that "by filling the hollow organs with opaque liquids, [such] as emulsions of bismuth in the stomach... their exact area can be readily determined."

In the years that followed, many other physicians used bismuth as a contrast medium to study the gastrointestinal tract. About that time the "bismuth meals" began to be replaced by barium sulfate ones. European radiologists had introduced the use of barium for their studies without any harmful effects, except when they used soluble salts of barium rather than barium sulfate. Barium was much cheaper than bismuth. Furthermore, when World War I broke out, American radiologists could no longer obtain bismuth from Europe, and barium became used almost universally. Barium sulfate is still one of the contrast media in general use by radiologists.

During the meeting of the American Roentgen Ray Society in 1908, papers were read on "X-ray Evidence in Gastric Ulcer," "The Roentgen Rays as an Aid in the Diagnosis of Carcinoma of the Stomach," and "The Roentgenographic Study of Motion in the Viscera." X rays were growing in importance to the diagnosis of disorders of the digestive system.

Lewis G. Cole
(1874-1954)

From Rigler, Leo G., and Weiner, Marvin: History of roentgenology of the gastrointestinal tract. In Margulis, Alexander R., and Burhenne, H. Joachim, editors: Alimentary Tract Roentgenology, ed. 2, St. Louis, 1973, The C. V. Mosby Co.

One remarkable radiologist specializing in the study of the gastrointestinal tract was Louis Gregory Cole. He was known as a maverick, but he also had a reputation of being extremely meticulous about every detail when he was making X-ray photographs. The radiographs he produced were known for their exceptional quality. He was described by some people as a "crank on the subject of immobility" while X-ray pictures were being taken. Cole not only insisted that patients be kept perfectly still (which no doubt contributed greatly to the clarity of the films) but he, like other radiologists of the era, had a collection of different X-ray tubes. He knew the characteristics and suitability of each one for different procedures. Henry Hulst wrote about how a radiologist regarded his tubes:

> He rests, pets, punishes, and smashes tubes as the fetishist his idols. He learns to know them all by name, their temper and their capabilities. He has a number of them—the more the better. He has trained tubes, trick

> tubes, high-spirited, high-bred tubes, as well as gentle, steady tubes.... A flashy tube is worse than an hysterical woman, it is incurably useless.

The recognition by Cole and others of the characteristics of the individual tubes gave an impetus to the development of improvements in X-ray tubes.

In 1914 while still in his thirties, Cole looked for signs of stomach cancers or ulcers on the X-ray films of 566 patients and made a number of correct positive diagnoses. Thirty-three of the patients on whom Cole made a negative diagnosis "presented sufficiently severe symptoms to justify surgical exploration"; 23 different surgeons operated on them. Remarkably, in all cases, surgery revealed that Cole's negative diagnoses had been correct. Cole had also proven how invaluable a diagnostic tool X rays could be to those who treated stomach and intestinal diseases.

The revolutionary effect of X rays on diagnosis of diseases and disorders of the gastrointestinal system has continued from these early discoveries to the present. Improvements in equipment and techniques have made it possible for physicians to diagnose and locate ulcers, cancers, anatomical abnormalities, trauma, circulatory problems, and a host of other conditions related to the digestive system.

The Brain

Another part of the body that held great interest for early roentgenologists was the brain. Edison was not the only one interested in photographing the living human brain. In the early 1900s, George Edward Pfahler became curious as to whether or not he could take an X-ray picture of a tumor in the brain. The problem he faced was that brain tumors are composed of soft tissues that differ only slightly in X-ray opacity, if at all, from normal brain tissues or from the surrounding fluids of the brain and spinal cord. Thus, most of the X-ray photographs of the brain examined by Pfahler did not allow the location of a tumor or other disease in the brain.

Pfahler discussed the matter with other physicians and scientists, including Arthur Willis Goodspeed, and found that there had been one distinguishable tumor demonstrated on an X-ray picture in 1899. Its presence had been confirmed on autopsy.

Pfahler continued thinking about using X rays to locate tumors in the brain. In 1901 he found a 32-year-old laundress who was willing to participate in his studies. Her right arm was paralyzed, and she had terrible headaches and other symptoms which suggested a large brain tumor. The laundress was placed in front of an X-ray tube, and the film was exposed for 4 minutes. The plate which had been placed 18 inches from the tube on the other side of the patient's skull "showed good detail of all the structures that might be expected in a normal brain; however, there was a large shadow lying between the coronal suture and the posterior meningeal artery." The shadow was shaped like a dumbbell, but was this the tumor or just an artifact (an artificial characteristic introduced by the technology)? For comparison, Pfahler took an X-ray picture of another person of similar size who did not have a brain tumor. That film did not show a shadow comparable to the one seen in the laundress's X-ray plate.

The surgeons decided to operate on the patient; when they did, the tumor was located in the exact spot where Pfahler had said it would be, but he was not pleased. The tumor the surgeons removed was much smaller than the one shown on his X-ray film, and it was not shaped like a dumbbell. "If I have demonstrated the tumor, then only half of the tumor has been removed," he said as the surgeons were completing their operation. Unfortunately, the seriously ill patient survived the operation for only a few hours. Pfahler attended her autopsy which revealed "a large remaining subcortical portion of the tumor...corresponding to the remainder and less definite part of the shadow."

George Pfahler (1874-1957)

From Rigler, Leo G., and Weiner, Marvin: History of roentgenology of the gastrointestinal tract. In Margulis, Alexander R., and Burhenne, H. Joachim, editors: Alimentary Tract Roentgenology, ed. 2, St. Louis, 1973, The C. V. Mosby Co.

Because of this initial accurate diagnosis, Pfahler was encouraged to perform many more experiments, primarily using the heads of cadavers. He improved the quality of his radiographs and gained confidence in his diagnoses for certain types of tumors. His later studies, however, showed that the great majority of tumors did not cast a shadow that would be visible on an X-ray plate. Therefore, other means were needed to identify these kinds of tumors.

One way in which the diagnosis of certain kinds of brain tumors was made possible was by deliberately introducing air into the ventricles (the fluid-filled cavities of the brain). Air provides a contrast to the soft tissues of the brain when viewed on radiographs, but it took several years before radiologists began using this technique.

A machinist living in New York City was the first patient in whom it was reported that parts of the brain had been outlined because the ventricles became filled with air. In 1912 he was hit by a trolley car and received an obvious head injury. The machinist was taken to Harlem Hospital where the radiologist, William H. Stewart, used X-ray films to diagnose a fracture of his skull. During the next 2 or 3 weeks, the man's condition grew worse, and a second set of X-ray photographs were taken. "To my surprise," Dr. Stewart reported, "I found we were dealing with a condition different than on the former examination. From the shape, location, and course of those shadows, their varying density and character simulating gas in the intestines, I was led to conclude that we were dealing with a case of fracture of the skull complicated by distended cerebral ventricles (inflated) with air or gas." One physician, Eugene W. Caldwell, to whom Stewart showed the X-ray pictures concurred that they were seeing pictures of air in the ventricles, but others doubted this diagnosis. After an operation on the machinist, the surgeon reported that "two or three quick spits of air and fluid" were found when he entered the cavity of the brain. The surgeon also found bubbles of air in the spinal fluid. The surgery, however, had come too late. During the autopsy of the patient a few days later when the brain was submerged in a tub of water and the ventricles were opened, air bubbled out. Apparently, in the time between when the first and the second group of X-ray photographs were taken, the patient had had a sneezing attack. The sneezing forced air through the fracture in the frontal sinus into the ventricles. This remarkable case was widely reported and discussed.

It might be expected that news of the case of the machinist would have prompted radiologists to begin introducing air into the ventricles deliberately for diagnostic purposes, but that did not happen for some time. During the next several years, many physicians undoubtedly saw sharp outlines of the brain because of air accidentally introduced into the ventricles, but no one took the creative step from accidental to deliberate introduction of air into the brain. This delay is even more surprising when one considers that the radiologists and neurosurgeons who must have viewed these plates were the same people who were most in need of a procedure to enhance visualization of the brain.

The person who finally did inject air to visualize the brain was a surgical resident at the Johns Hopkins Hospital. For some time Walter E. Dandy had been interested in methods of diagnosing brain tumors. At the time only tumors which contained calcium could be seen on an ordinary X-ray photograph. Calcium is opaque to X rays, but the process of tumor calcification rarely occurs to the extent that it would have been visible on the X-ray pictures of that era.

Dandy and George Heuer, his associate at Harvard, reviewed X-ray photographs of 100 patients in whom the existence of brain tumors had been demonstrated by surgery. Only six of these had a shadow that might have indicated the tumor before the operation. In another nine cases, the tumors were obvious because they distorted the cavities of the brain. The remaining 85 tumors could not be seen on X-ray photographs. To obtain better diagnostic accuracy in locating brain tumors, Dandy and Heuer decided that, if they could fill the ventricles with some substance which would produce a shadow, they could find a number of the other tumors. Many tumors change the configuration of the walls of the ventricles. If a change in their shape could be detected, it would be a clue that a tumor was present. Dandy and Heuer decided that whatever they injected had to be nontoxic and readily absorbed and excreted by the body. They tried many substances (including salts of bismuth, potassium, and thorium) on laboratory animals, but always the substance caused fatal injury to the brain. Eventually, they decided it was unlikely that they could find a substance safe enough to inject into the brain. They were both unaware of any of the several published reports about air accidentally introduced into the ventricles.

Dandy's insight came in a very unusual way through his surgical chief, William S. Halsted. It seems that when Halsted and his staff were reviewing X-ray photographs that showed bubbles of gas in the intestines, the bubbles were recognizable even if a layer of bone were between them and the X-ray plate. This circumstance sometimes prompted humorous remarks from Halsted. In his jokes, he often referred to the way in which intestinal gas had the ability "to perforate bone." Later, when Dandy wrote about his moment of insight, he said that his "attention was drawn to [the] practical possibilities [of introducing gas] into the brain." It had occurred to Dandy that, if the stomach or intestines could be seen more clearly when they had air in them than after they had been filled by a bismuth meal, the radiographic properties of air might be just what he needed to solve his problem.

One might think that a procedure so unusual would be tried out first on laboratory animals; however, no animal experiments were reported. By July of 1918, Dandy had published an account of the first 20 human cases in which air had been introduced into the ventricles for diagnostic purposes. He found that "air and water in a ventricle behaved exactly as they would in a closed flask. Following any change in position, fluid gravitates to the most dependent part and the air rises to the top." Dandy was thus able to position an air bubble anywhere in the ventricular system he wanted it by positioning the patient's head in different ways. His experiments led the way to many improvements in studying the brain and spinal cord, and Dandy did pioneering work using air to visualize other parts of the cerebrospinal system.

The Cardiovascular System

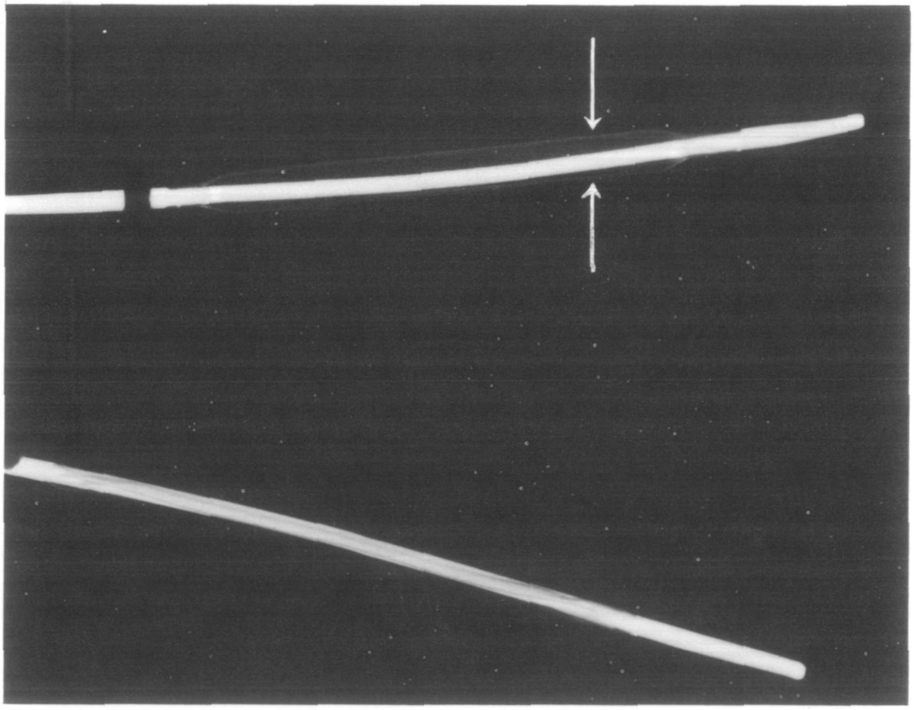
Physicians also were fascinated with the idea of studying the circulatory system of the living human body, but it was not safe to use air, bismuth, or barium for this purpose. A nontoxic substance that was opaque to X rays, soluble in the bloodstream, and easily absorbed by the body was needed. The first compound which satisfied these criteria was potassium iodide. In 1919 its use was first reported by Carlos Heuser, an Argentinian radiologist. Unfortunately, his finding received little recognition, probably because his account was written only in a Spanish-language journal.

This use of potassium iodide and other compounds (such as sodium iodide) in the same chemical family, had to be found all over again. At the Mayo Clinic, Earl Osborne was studying the effects on the body of sodium iodide injections which were used to treat patients suffering from syphilis. He demonstrated that when sodium iodide was introduced into the bloodstream, it was concentrated by the kidneys and then excreted. Leonard Rowntree, who was also working at the Mayo Clinic and who was interested in diagnosing kidney diseases, knew that sodium iodide was opaque to X rays. He concluded that X-ray photographs might be taken of the kidneys without having the difficulties associated with injecting substances such as barium into the ureters (the tubes which connect the kidneys to the bladder) by way of a catheter through the bladder. His idea was an easy one to test because patients were already receiving sodium iodide injections to treat syphilis, and many of those same patients volunteered to have X-ray photographs made. The doctors found that by using this method they could now visualize kidneys as well as the rest of the urinary tract by using X rays.

Osborne's and Rowntree's work with sodium iodide was the basis for new work to visualize blood vessels. They had noted that, when the fluoroscope was used to watch the injection of sodium iodide into a vein, that vein had "the appearance of a steel wire" because of the iodide. In this instance, several of those who were aware of this finding immediately saw its potential value in visualizing blood vessels in the living human body. A rapid series of experiments proved them to be correct.

These are just a few of the numerous examples of discoveries in the early days of diagnostic radiology which were the foundation of the very sophisticated techniques for injecting radiopaque substances into various parts of the cardiovascular system. Because of these key discoveries and many subsequent ones, doctors now are able to perform diagnostic procedures called angiography. These studies involve passing a small plastic tube (a catheter) through veins or arteries. The catheter is guided to the desired location in the body while the physician watches on a fluoroscope to determine the exact location of the catheter's tip. When the catheter is properly positioned, liquid which is opaque to X rays is injected into the catheter and thus into the blood stream. This material temporarily replaces the blood flowing through the area and outlines its arteries, capillaries, and veins. The angiograph (an X-ray photograph of the blood vessels containing the opaque substance) gives a physician important information about whether or not there is abnormal anatomy or disease in the region under study.

Improved fluoroscopic technology now allows radiologists to inject a substance visible on X-ray photographs into a superficial vein in the arm. They can then observe this substance as it passes through the body with the normal blood flow. Thus, problems in certain large arteries (such as those in the neck) can be identified at relatively little risk or discomfort to the patient.



An inflated balloon-tip catheter such as the one above may be used to open narrowed or blocked blood vessels. The lower catheter is not inflated.

David O. Davis

Using catherization procedures, physicians frequently can inject substances which stop vessels from abnormal bleeding. This technique is not only beneficial because major surgery can sometimes be avoided but is also effective in some cases in which surgery is impossible.

When important arteries in the body become narrowed, a catheter can be inserted to the site of the problem. The blood vessel is then dilated by blowing up a tiny balloon at the tip of the catheter. When the balloon is deflated and the catheter is removed, the blood can once again flow through the vessel, but this time without obstruction. This method is now also being used in selected cases where arteries of the heart have become narrowed. The long-term effects of this procedure are not known, but there is evidence that the balloon-dilated vessels remain open.

These new methods of blocking bleeding vessels and opening obstructed ones are the basis of a new subspecialty in the field of radiology called interventional radiology. The goal of interventional radiology is to eliminate the need for major surgery in some patients who previously had no alternative.

DIAGNOSTIC X-RAY EQUIPMENT

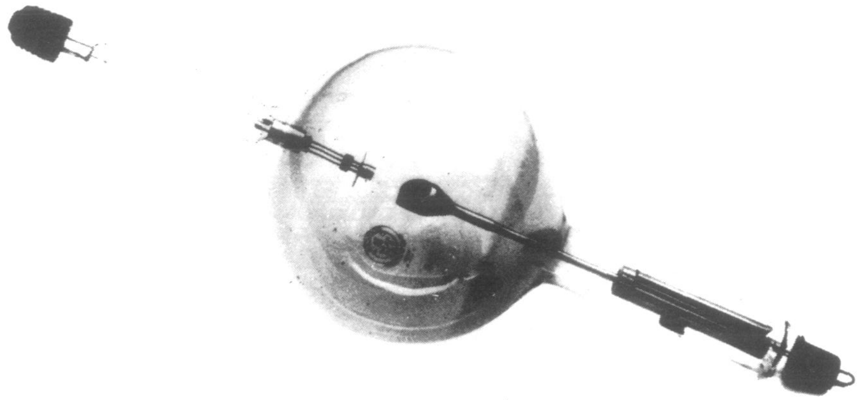
Many of the advances in diagnostic radiology would not have been possible without improvements in X-ray equipment.

Coolidge Tube

Conventional Crookes tubes had many limitations. They would not work unless a residue of gas were left inside them. Problems occurred because the properties of the gas in the tube changed when the gas was heated, and these characteristics varied from day to day and from minute to minute.

While studying at Massachusetts Institute of Technology in 1896, William David Coolidge did not have his fascination with X rays diminished when he had to be treated for X-ray burns; however, he did gain a great respect for the rays. He was also interested in making better electric light bulbs by developing filaments which were less brittle. He developed a process to improve the tungsten that was then being used to make the filaments; he called his new product ductile tungsten.

For months Coolidge made efforts to improve Crookes tubes by using ductile tungsten. When an ordinary tube was activated there was an electrical potential between the cathode and the anode. This caused positive ions of gas in the tube to strike the cathode. This bombardment brought about emission of electrons which then struck the



The Coolidge tube was introduced in 1913. It provided for the first time, the ability to determine and control radiation precisely.

from "The Trail of Invisible Light" by E.R.N. Grigg, M.D.

anode and thus generated X rays. The problem was that if there were not enough gas in the tube to make ion bombardment of the cathode possible, the tube did not work. Thus the problem of producing a stable tube by decreasing the amount of gas in it did not seem to have a solution unless the electrons to hit the target could be supplied in another way.

An Austrian researcher, Julius Edgar Lilienfeld, perceived the same problem that Coolidge did. Lilienfeld solved the dilemma by building a tube containing a filament which, when heated, became incandescent and produced electrons. These electrons made it possible to produce X rays, even though almost all of the gas had been evacuated from the tube. Lilienfeld applied for a patent that specified that electrons from the hot filament were used to lower the resistance of the tube.

A number of other people were working on making a better X-ray tube, but the tube designed by Coolidge in 1913 having a pressure "as low as it has been possible to make" was the breakthrough. In his first experiments with his new tube, he recalled, "I temporarily and unintentionally sacrificed my own back hair. So I didn't like to practice on other living subjects. For further experiments I was, through the kindness of a medical friend, provided with a human leg which had outlived its usefulness."

Coolidge made some initial tests on the cadaver's leg with an innovative tube. His invention had a high vacuum, a hot cathode, and a tungsten target. He then gave his tube to Louis Gregory Cole, the same New York radiologist whose contributions to gastrointestinal radiology already have been described. Cole described Coolidge's tube

William Coolidge (1873-1975)

as "undoubtedly the most important contribution to Roentgenology since the birth of that science." Cole published material which proclaimed the advantages of the new tube. With the Coolidge tube radiologists could expect an exact duplication of their previous results, and furthermore, the tubes could be adjusted very accurately. Because he was known to have such high standards for his own work, praise by Cole of such an invention promoted the tube almost overnight. The General Electric Company was flooded with orders for the tube from almost all of the radiologists in the United States, and the tube became standard equipment for radiologists everywhere. In principle, the Coolidge tube was also the prototype of the electron tubes used in radios and computers and for many other purposes, until their gradual replacement by transistors and solid-state devices.

There were numerous other advances in the equipment that was used by radiologists. Power supplies were improved; and, with the coming of World War I and the subsequent unavailability of European glass, the glass photographic plates which were used with the overwhelming majority of the X-ray equipment were quickly replaced by film. The early film, made of cellulose nitrate, was very flammable and gave out large quantities of poisonous gas when it burned. In 1929 more than 100 people died in a fire in Cleveland caused by flammable film, an incident which helped make people aware of the hazard. Improved cellulose acetate film, which was available by then, was quickly adopted and is still in use.

Image Intensifier

Technologies such as the Coolidge tube and improved X-ray films were critical to the progress made in diagnostic radiology. For about 50 years after the discovery of X rays, radiologists had seen the inside of patients by directly examining a dimly lit fluorescent screen which had been excited by X rays after passing through the patient. The image was so dim that radiologists routinely had to spend about 20 minutes adapting their eyes to the dark so they could see the image on the screen.

In 1928 J. H. deBoer, G. Host, and M. C. Teves of Philips Corporation in the Netherlands filed for a patent on the first device that formed a brighter image by using the electrons from the fluorescent screen. However, they encountered what were at the time insurmountable technical difficulties and were able to achieve only a modest laboratory model.

V. K. Zworykin and G. A. Morton of the Radio Corporation of

America Research Laboratories published their work in 1934 on a similar tube which they had developed to convert invisible infrared or ultraviolet images to visible light. During World War II an intense effort was mounted in Germany and in the United States to apply these techniques towards the development of a tube, illuminated with infrared light, which could be used in sniperscopes and snooperscopes for observing the enemy at night.

It was not until 1948, however, that John W. Coltman, a physicist at Westinghouse Research Laboratories, published the report of his successful development of the first image intensifier. This device succeeded in converting an X-ray image into a visible image as much as 500 times as bright as those previously obtained. Modern X-ray tubes employ the basic principles of Coltman's tube, but they provide more than 5,000 times the brightness of the early standard fluorescent screens.

The invention of the X-ray image intensifier revolutionized the practice of diagnostic radiology. Almost immediately the intensifier was used to observe the heart in motion. At last there was enough light to expose movie film, and the indispensable technique of cinefluorography was born. The developments in television technology in the 1950s made it possible to introduce closed circuit television monitoring of fluoroscopy procedures. Improvement in, and simplification of, television equipment in the 1960s brought about the widespread use of television monitors. Then video tape recorders made it possible to record complete fluoroscopic procedures in real time and also in playback and slow motion for more accurate diagnosis.

But even in the 1960s, the quality of images obtained with image intensifiers was not as good as the quality of images on radiographic film—primarily because the substance on the intensifier screen, zinc sulfide powder, was not dense enough to give a sharp image. Furthermore, when light struck the screen it scattered from one grain of powder to the next, thereby causing "smearing" of the image. Eventually, a tube was developed which used cesium iodide instead. This substance gave a surprisingly good image. Research showed that when cesium iodide was applied to coat a screen (instead of depositing as a smooth, continuous layer as expected), it deposited as a layer of material that could be conceived of as millions of closely packed needles, all parallel to each other and standing up from the sheet of glass. Each needle acted as an optical fiber conductor of the light produced in it by the X rays. (Optical fibers are discussed in more detail in the chapter, "Synthetic Fibers.") The "needles" efficiently conducted the light to the surface of the instrument without scattering it sideways to neighboring needles. Thus, the image was much sharper than ever before.

Tomography

A conventional radiograph is obtained by passing a beam of X rays through the body, and capturing the images projected by the X rays on two-dimensional film. But when X rays pass through a patient, dense tissue such as bone obscures less dense tissue which may lie in front of or behind it. Thus, only structures not in line with such dense tissues can be visualized easily. A device was needed which would make it possible to "photograph" parts of the body without those parts being obscured by the shadows of intervening tissues. Today such machines are called tomographs, and the pictures they produce are tomograms.

Perhaps Carol Mayer, a Polish radiologist, took the first step towards developing the first tomograph. Apparently in trying to get better X-ray pictures of a patient's heart, he moved his X-ray tube. The movement of the tube blurred the shadows produced by the patient's ribs, which were the part of the body closest to the tube; but the heart, which was further from the tube, had a better defined image than the ribs did. He published his findings but did not apply for a patent.

Jean Kieffer (1897-)

The work of Mayer probably was not known to Jean Kieffer, a young man afflicted with tuberculosis. Kieffer was an X-ray technician who worked in a tuberculosis sanitarium in Connecticut following his own trials with tuberculosis. While confined to bed for treatment, Jean Kieffer used the time to think of a way to visualize the region of the body, the chest, where his own disease was concentrated. The critical part of his invention was an X-ray source placed a fixed distance from an X-ray film. The source and the film could be manipulated around a pivotal point or fulcrum—somewhat like two ends of a beam on a balance scale. The device could be moved in broad arcs, shallow arcs, or in complex circular, curving, or spiral patterns. During the movement, both the X-ray tube and the X-ray source remained in a proportional relationship to the pivotal point, thereby making it possible to keep an object in the plane of the pivotal point in focus on the X-ray film. Most of the shadows of objects which were not in that plane were "erased." Kieffer also found that as the point of rotation was moved closer to or farther from the film, different planes could be brought into focus. He applied for a patent on his "X-ray focusing machine" in 1929.

For the next several years Kieffer tried to interest manufacturers in building his machine. Because he had not completed high school and his invention was supposed to be the solution to a problem the manufacturers themselves had not been able to solve, he could not find anyone who would give financial backing for his device.

He must have had mixed feelings when he read the headlines of an article in the *New York Times* of September 29, 1936, which said, "New X-ray Device 'Dissects' by Films—Machine Makes Possible Photographs of Parts of Organs Unobscured by Tissue." J. Robert Andrews and Robert J. Stava of the Cleveland University Hospitals had perfected a tomograph and demonstrated it to those who had arrived early for the meeting of the American Roentgen Ray Society. Kieffer immediately left for Cleveland to attend the opening day of the meeting, which was scheduled for the following day. Jean Kieffer still had an important part to play in the history of tomography.

Most of the radiologists at the Roentgen Ray Society meeting were not impressed by Andrews and Stava's discovery. The films which were exhibited were considered very poor from a technical point of view. Less detail seemed to be visible than in standard X-ray photographs.

In addition to Kieffer, another man at the meeting had tremendous interest in Andrews and Stava's exhibit and saw great potential in the discovery. He was Sherwood Moore, director of the Edward Mallinckrodt Institute of Radiology at Washington University School of Medicine in St. Louis. He struck up a conversation with Andrews and Kieffer (who also had come to visit the exhibit booth). Moore particularly was impressed as he heard Kieffer describe his own ideas about a similar machine.

Moore wanted Kieffer to help him supervise the building of such a machine in St. Louis, and Kieffer took a leave of absence from his job to do so. In a 1937 letter written home to his wife, Kieffer gave a candid description of his feelings:

> I am so tickled at the machine and what it does that I would like to dance a jig! Dr. Moore and I almost did that yesterday morning when I got a film that he didn't think could be gotten... a picture of an abdominal aneurysm! The impossible! And it was also of very much importance because it absolutely proved the diagnosis, which was questioned by some men. Boy—Oh—Boy! I was working on the machine when Dr. Moore came up with the film as it came out of the dark room. "Come see this!" he literally yelled...It definitely showed things you wouldn't guess with regular X-ray film...

When Dr. Moore saw the long motion working [you know, he was somewhat skeptical], he said, "I won't believe it until I see it again," and he did. He was all smiles and shook my hand and said, "Kieffer, we've got them all licked."

COMPUTED TOMOGRAPHY

The development of X-ray tomography was an important step in improving the physician's ability to visualize parts of the body without their being obscured by overlying tissues; however, there were other limitations of conventional X-ray technology yet to be overcome. Ordinary X-ray films can differentiate between air, bone, fat, and water densities but not between those of certain soft tissues which have densities that differ from one another less than 2½ percent. Moreover, many structures in the body can be visualized only by introducing contrast media such as air or substances which are opaque to X rays. Variations within soft tissues such as the brain, liver, spleen, and pancreas are usually not distinguishable, and conventional X-ray photographs cannot be used to measure accurately the differences between objects with similar densities. The development of the technology of *computed tomography* (CT) has made it possible to reduce these limitations.

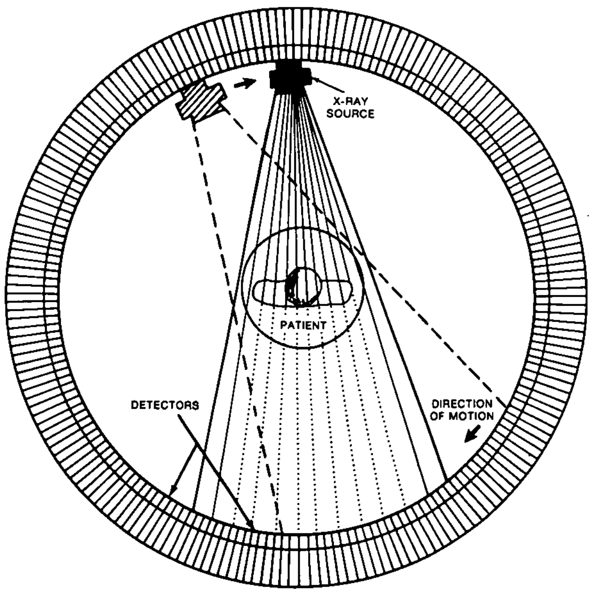


Diagram of one type of CT Scanner

With computed tomography a radiologist can obtain clear representation of a desired slice of the body without the blurred-out anatomical detail from unwanted intervening structures—a characteristic of conventional tomograms. In a CT scanner, an array of X-ray detectors is situated opposite the X-ray tube on a large metal ring. A patient lies on a table that can be positioned through the center of the ring. The table is then placed so the particular area of the body to be examined is between the X-ray source and the detectors. As X rays are produced, the ring rotates. When the X-ray source has rotated in a complete circle, the scan of that slice is complete. At the same time the ring is rotating, the detector translates the measurements of the changes in the X-ray beams as they pass through the object into electrical impulses which are then transmitted to the computer. These impulses are quantified, mathematically manipulated, and then used to reconstruct a tomographic image.

Computed tomography is the "reconstruction" of an object through a particular plane (a "slice") using X rays and a computer, and the machinery developed to perform these procedures is referred to as a CT scanner. Another term frequently heard is *computerized axial tomography* (CAT). The latter term is more restrictive in meaning because it refers to computer reconstruction of cross-sectional images that essentially are transverse to the long axis of the patient, whereas "CT" refers to reconstruction of a slice through the body taken at any angle. An even more general term is *computerized reconstruction*. This terminology may be applied to nonmedical as well as medical uses of the technique and to reconstruction in either two or three dimensions.

One key to the development of computed tomography was finding a way to reconstruct the internal structure of an object from X-ray measurements made external to the object. As early as 1917 an Austrian mathematician, Johann Radon, published his investigations which ultimately were applicable to the theory of reconstruction. But Radon was not working on image reconstruction; he was solving gravitational field equations. Perhaps because his research was in a different field is why his work was not known to some of the pioneers in computed tomography who had to develop their own equations.

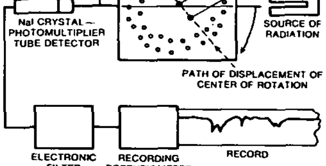
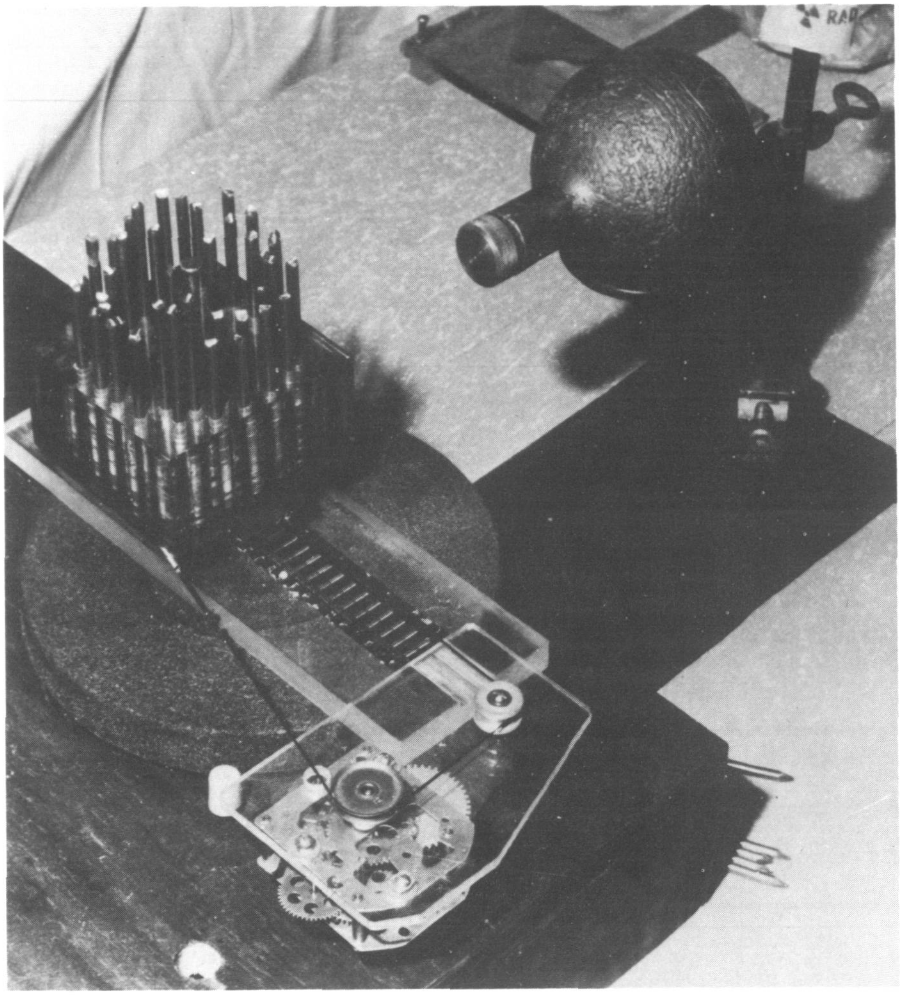


Diagram of Oldendorf's Experiment

In 1956 a neurologist, William H. Oldendorf, was the first to attempt a medical application of image reconstruction. He wanted to find a way to make diagnostic X-ray images of the brain which were not complicated by obstruction from the overlying bones of the skull. Oldendorf made a small block of plastic and inserted iron and aluminum nails into it. He then put the block on a toy train "HO" gauge flat

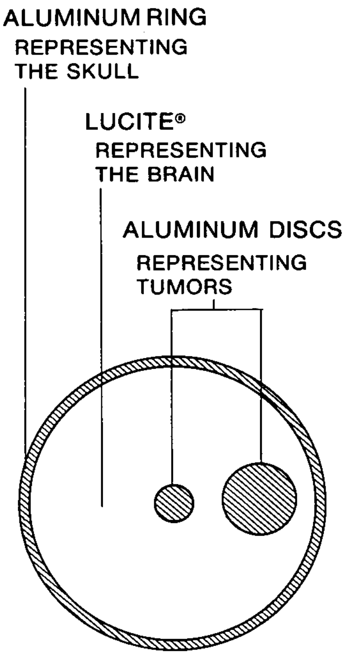


This photograph shows Oldendorf's experimental apparatus.
William Oldendorf

car which was pulled very slowly along a short track by a clock motor. The whole assembly was mounted on a phonograph turntable which rotated at 16 times per minute. Gamma rays passed through the center of rotation of the turntable, and the plastic block was pulled through this center of rotation. Using this apparatus Oldendorf was able to obtain an approximation to his model.

Another critical link in the development of computed tomography was the work Allan M. Cormack published in the early 1960s. The story of Cormack's contribution to CT scanning began in 1956 when he was a lecturer in physics at the University of Cape Town in South Africa. In that year the staff of Groote Schuur Hospital unexpectedly found itself in need of a qualified physicist to supervise the use of radioactive isotopes. Because Cormack was the only nuclear physicist in Cape Town and the presence of one was required by law, he was asked to spend 1½ days each week at the hospital to fulfill this stipulation. While working at the hospital he became familiar with the planning of radiotherapy treatments and was intrigued by the complications of calculating the appropriate doses of X rays for patients. In his own words:

> It occurred to me that in order to improve treatment planning one had to know the distribution of the attenuation coefficient of tissues of the body, and that this distribution had to be found by measurements made external to the body. It soon occured to me that this information would be useful for diagnostic purposes and would constitute a tomogram or series of tomograms, though I did not learn the word tomogram for many years.



Cross-sectional Diagram of Cormack's Phantom in 1963.

Cormack knew that the basic research on attenuation of X rays in homogenous substances had been done nearly 60 years earlier, but when he searched the literature for reports about what happened when X rays passed through substances which were not homogenous, he found nothing. Because he did not learn until 14 years later that Radon had solved the problem in 1917, Cormack started his research without this basic information.

One of Cormack's first experiments was done on a symmetrical phantom—a cylinder of aluminum surrounded by a wooden ring, diagramed on left. (Phantoms are nonliving experimental substitutes for patients which can provide information, usually quantitative, about how a particular machine will perform with a living subject.) Because the form was symmetrical, fewer measurements were needed than for an asymmetrical form. Gamma rays were projected from a cobalt source through the phantom, and a Geiger counter was used to detect the radiation which passed through the phantom. The data were processed by laborious hand calculations, and when they were analyzed they agreed with the known structure of the phantom.

Cormack continued his work intermittently over the next 6 years. In an experiment in 1963, his phantom was more complicated—it was an outer ring of aluminum representing the skull with Lucite® inside representing the soft tissue of the brain. Two aluminum discs representing tumors were placed inside the Lucite®. Cormack used a computer to manipulate data (obtained by measuring the radiation that passed through the phantom) that it had taken 2 days to collect. The results from Cormack's first computed tomography agreed with what would be expected from the known form of the phantom.

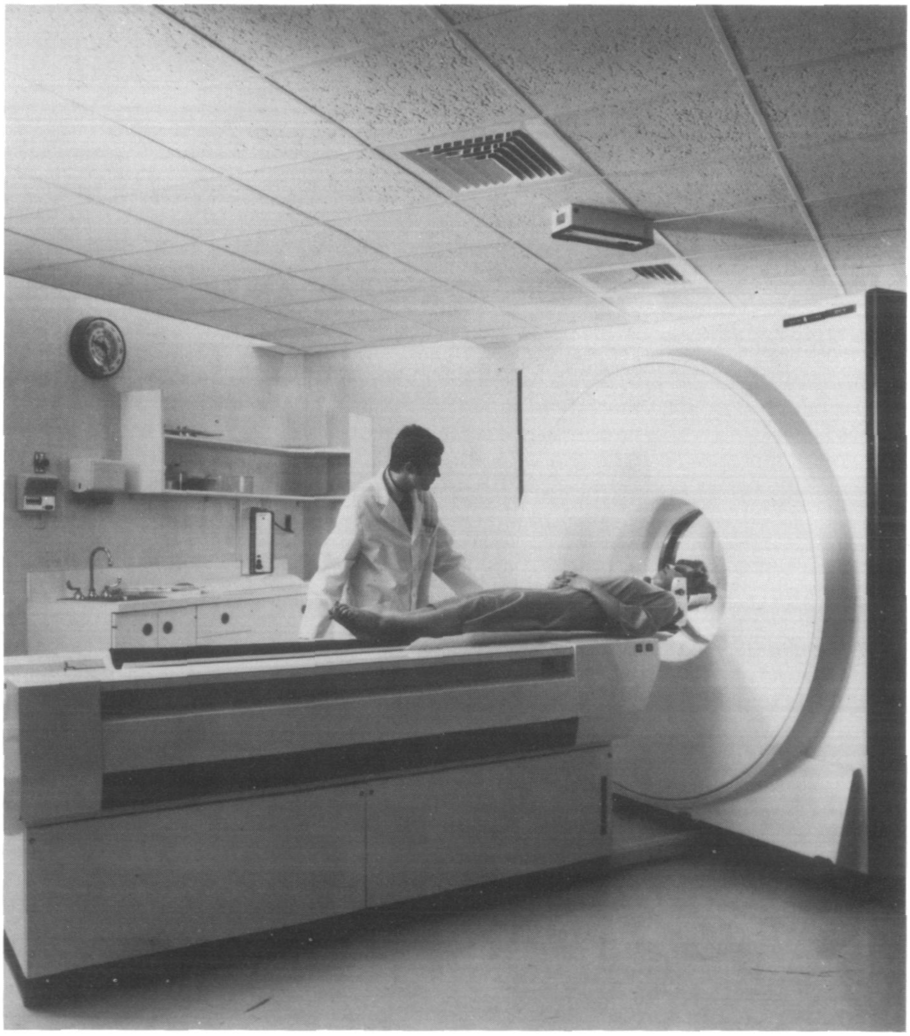
Cormack's work was published in 1963 and 1964 and went almost unnoticed except by a few scientists. About 10 years later, and totally independently, the first clinically useful computed tomography equipment was developed. It was primarily the result of the work of Godfrey N. Hounsfield who was working at the Central Research Laboratories of Electrical Musical Instruments, Ltd. (EMI) in England. (EMI is a major force in the world's music business and the producer of the Beatles' records, among others.) Hounsfield's research had been in computer techniques for pattern recognition of characters written on business forms. With this background and his interest in information retrieval in conventional X-ray examinations, Hounsfield undertook gamma and X-ray experiments in tomographic reconstruction. He speculated about the possibility of using a computer to reconstruct a picture from measurements of X rays that had passed through the body at many different angles. Because of the multitude of measurements that would have to be made before a picture could be reconstructed from them and the great number of simultaneous equations which would have to be solved, he knew a computer was essential.

Hounsfield set out to see if he could develop a method to distinguish between tissues having very similar densities. In his words, "The equipment was very much improvised." A lathe bed provided the lateral scanning movement of the gamma-ray source, and sensitive detectors were placed on either side of the object to be viewed, which was rotated 1° at the end of each sweep. The 28,000 measurements from the detector were translated into numbers and automatically recorded on paper tape. After the scan had been completed, the results were fed into the computer and processed.

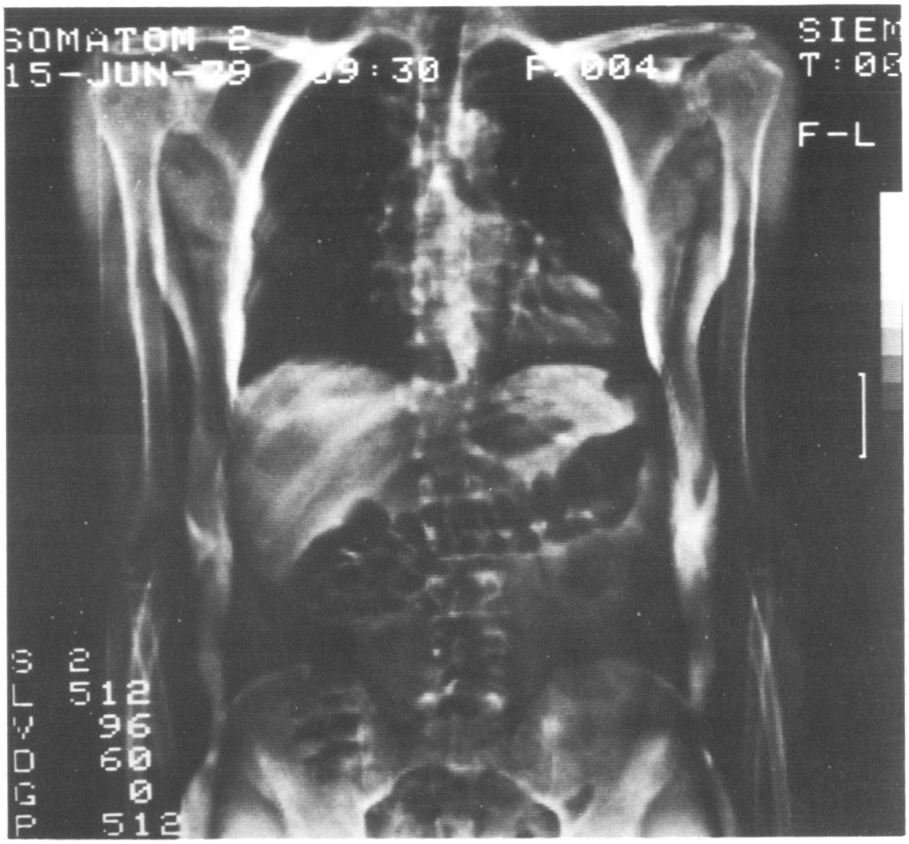
Hounsfield's machine produced encouraging pictures, but it worked quite slowly. It took 9 days to obtain the data by projecting X rays through a phantom and 2 ½ hours to process the results on a computer. This method obviously involved using too much radiation and was too tedious and time-consuming to be of clinical use. The next technological advance was the installation of a more powerful X-ray source, which reduced the scanning time to 9 hours. With the help of two radiologists, James Ambrose and Louis Kreel, Hounsfield used preserved human brains and unpreserved animal brains to test the equipment. He and his associates were successful in reconstructing brain tissue, but the crucial test was whether or not tumors could be detected.

Eventually, faster and more sophisticated machines were built to scan brains of living patients, and in 1972 a woman with a suspected brain lesion was examined using a machine developed by Hounsfield. The picture revealed a circular cyst in the woman's brain. As Hounsfield put it, "From this moment on, as more patients were being scanned, it became evident that the machine was going to be sensitive enough to distinguish the difference between normal and diseased tissue."

The total time elapsed from Hounsfield's idea to the production of a functioning clinical machine was about 5 years. The total investment in research and development during that time was about $5 million. In the less than 10 years since the invention of CT scanning, about 100 times this amount has been invested in research and development of CT scanners by industrial firms in the United States, Europe, and Asia.



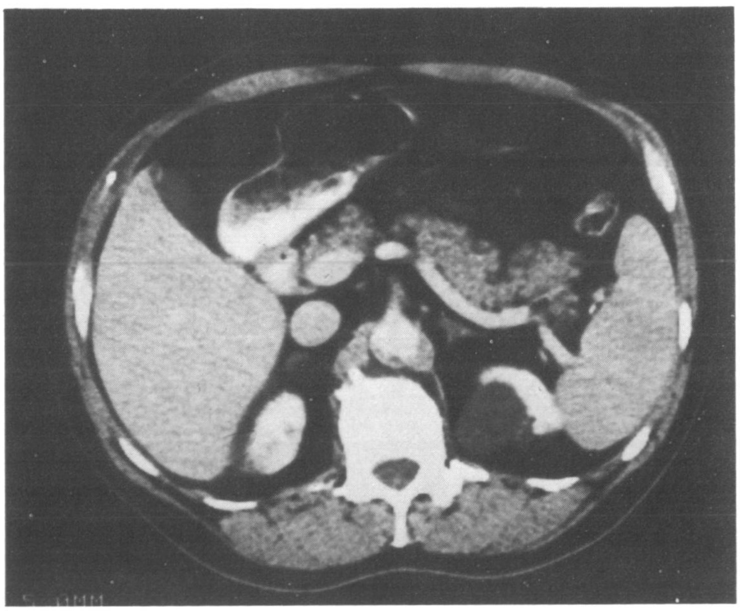
This highly advanced CT system is capable of completing an entire scan of the head or body in as little as 4.8 seconds, thereby minimizing image distortion caused by patient movement.
General Electric Company



This image (above) resulted from a preliminary CT scan of the thorax and abdomen. From views similar to this, physicians can decide precisely where they want to make more detailed cross-sectional CT scans such as the one shown (below) of the upper abdomen.

Siemens Corporation

General Electric Company



The first use and advantage of the CT scanner over conventional X-ray methods was in visualizing the brain, but its value has rapidly been proven in other regions of the body.

There have been many improvements in CT scanning since its discovery. The first EMI machine cost about $300,000 in 1972; in 1980 a CT scanner with similar capabilities can be purchased for about $120,000. However, some more sophisticated instruments now may cost up to $1,200,000, and operating costs for a unit have been estimated to range from $259,000 to $371,000 per year. The initial estimate of the world-wide market for CT scanners was for less than two dozen machines. Now the number of scanners in use is well over 2,000, and new ones are being manufactured each year.

Such high costs for a single piece of medical equipment have brought CT scanners very much into the public eye, particularly at a time when so much attention is being focused on the skyrocketing costs of medical care. The rapid acceptance of CT scanning has also promoted concern over the cost of other diagnostic technologies and even medical technology in general.

Many policy issues related to CT scanning have arisen: How and by whom should the need for and placement of CT scanners be determined? When is a scanner safe and efficacious? Is its cost justified? Who should pay for the technology and its operation? Some of these issues have been resolved, and considerable effort is continuing at national, state, and local levels to find answers to other related policy questions.

The field of computed tomography is still developing. Improved resolution and shorter scan times are being sought to examine not only the brain but also other parts of the body. Currently, scientists, engineers, and clinicians are trying to obtain useful diagnostic pictures of the heart using CT; other problems, such as motion, must still be solved.

In 1979 Allan Cormack and Godfrey Hounsfield were awarded the Nobel Prize in Physiology and Medicine for their contributions to the development of computed tomography. Neither man had a background in physiology or in medicine. The invention of CT is one good example of the dependence of technological developments on previous discoveries in many, diverse scientific fields. In this case, contributions from the disciplines of mathematics, physics, computer science, medicine, physiology, and engineering were all brought together to develop a technology which probably has had the greatest effect on the field of radiology in the 85 years since Wilhelm Conrad Roentgen saw that mysterious glow in his darkened laboratory.

Acknowledgements

In addition to those mentioned in the Preface, the following people have contributed to the development of the Twelfth Annual Report of the National Science Board by providing expert suggestions, expert advice, information, documentation, photographs, written material, or technical assistance. Their contributions are greatly appreciated.

Bruce R. Abell
Murray Aborn
Marilyn Abrams
H. Adams, Jr.
Kathy Adgate
Dino Agro
Carrie C. Akin
John R. Akin
Margaret M. Akin
Laura Ash
James Atkins
Milo M. Backus
Kathy Bacon
William O. Baker
Albert Barber
Keith C. Barrons
Victoria Berdon
Walter E. Berdon
Norbert M. Bikales
Robert Bjork
Michele Black
J. R. Blackwell
Etcyl H. Blair
William A. Blanpied
John A. Blume
Jaan Born
Albert Bowers
Mary Boyd
William Boyd
Robin P. Brett
Harvey Brooks
Reynold F. Brown
H. Joachim Burhenne
Arthur W. Burks
Glenna Burns
Ann L. Burroughs
David N. Byrne
Albert H. Cantril
Susan Cantril
Barbara Carroll
John E. Casida
Donald Chiappenelli
Elizabeth Clark
Gwen Clopton
Curtis Coker
Cynthia Conroy
Jean Converse
Allan Cormack
Franklin Cosmen
Richard Coyle
Bryce Crawford
Barbara J. Culliton
Kent Curtis
John Davidson
David O. Davis
Milton Dobrin
James Donahugh
Laurel S. Donovan
John Doppman
Richard Drew
Joan Durgin
James Early
Robert Eilers
Stanley F. Englebardt
Brian Erb
Gerald Estrin
Ronald Evens
Gail A. Fachetti
Patricia Failla
Coralie Farlee
W. Geary Flamm
Catherine Flynn
William Foege
Virgil Freed
Richard J. Fox
Donald Frenzen
Gordon M. Frey
Shigeru Fukuda
Denos Gazis
James F. Gibbons
John H. Gibbons
Stephen A. Gilford
Roger L. Gilkeson
Edward L. Ginzton
Edward Glass
Henry I. Goldberg
Derek V. Goodwin
Cleve Goring
Jere E. Goyan
J. Thomas Grayston
Adele Greene
Thomas Griffin
Hans Grillmeyer
Lawrence D. Grouse
Carol Hallman
Joyce Hamaty
Milton Harris
David Harrison
Wayland J. Hayes, Jr.
J. C. Headley
Hollis Hedberg
Mark Hegstet
Charles H. Herz
William R. Hewlett
Lawrence Heyl
Ira M. Heyman
Roger W. Heyns
Richard N. Hill
Harman Hoffman
Norman R. S. Hollies
Gerald Holton

Barbara Hopkins
George L. Hutton
Herbert Hyman
Edward Hymoff
D. M. Jacobsohn
Ursula Johnsen
Hamilton Johnson
Julius Johnson
Raymond Johnson
Paul Jolly
J. Arthur Jones
L. B. Jones
Walter J. Karplus
Robert Kates
James B. Kendrick
William Kerr
Lucinda Kiester
Charles Kittel
Lisa Klein
Nancy Knight
Donald Krotz
John Kucharski
Arthur Kusinski
Roger O. Lambson
Kenneth L. Larner
Otto Larsen
Marilyn Laurie
David Leege
Martin Lefcowitz
Donald Lerch
Franklyn Levin
Otha W. Linton
John G. Linvill
Mary Ann Lloyd
John A. Lofgren
John L. Lundberg
Richard W. Lyman
Barbara S. Lynch
Cynthia A. Lynch
Stuart Lynn
Paul Lyons
William D. McElroy
Donald H. McLaughlin
Thomas Malone
Betty W. Maloney
Alexander R. Margulis
Herman Mark
Kent Marquis
Archer Martin
L. John Martin
George J. Maslach
Harry Mayne
Robert L. Metcalf
Michael Mikuda
Thomas Milby
Earl R. Miller
William Miller
O. B. Mobley, Jr.
Paul Morgan
Russell Morgan
Alexander J. Morin
Audrey Morrill
Carl Moskowitz
Emil M. Mrak
Robert Mulholland
William S. Murray
Richard P. Neil
Daniel Newman
E. John Northwood
Craig Nunan
William H. Oldendorf
Geoffrey Oldham
Gilbert Omenn
D. P. O'Neil
Jane C. Orr
Mark Ouyang
Ronald Overmann
Walter Parham
Mary Parramore
Betty Partin
Louise Patenge
J. K. Perrell
Nicholas L. Petrakis
Paul D. Phillips
Theodore L. Phillips
Frances A. Pitlick
Willa Polk
Phillip W. Porter
Kenneth Prewitt
Judy Price
David Prowitt
Mary Ann Rabin
Ludwig Rebenfeld
Leonard A. Redecke
Robert Reinke
Robert B. Rice
Phillip C. Ritterbush
John P. Robinson
Sadie R. Robinson
Joe J. Robnette, Jr.
Joseph V. Rodricks
Joanne R. Rom
Shirley Rosenberg
Richard Royston
F. James Rutherford
Adaline Ryan
James Schmiedeskamp
Aaron Seriff
Louise W. Sessions
Nathaniel Sheppard
Robert E. Sheriff
Ray Shockley
William B. Shockley
Nancy Siegal
Elliot Sivowitch
Leslie Sklar
Philip Smith
Ray F. Smith
Sukari B. Smith
Edward A. Smuckler
Mason P. Southworth
James Stewart
Clare Stober
George Streisinger
Arthur M. Sussman
Millard R. Swingle
Bernie Talbot
David Torgerson
Darleen Treneer
Solomon Triebwasser
Earl Ubell
Bruce L. Umminger
William M. Upholt
Marcia Van Etten
Ian Vingoe
Robert Vogel
Rose Volkert
Sheldon L. Wagner
Wallace K. Waterfall
William Weiss
Bonnie Weissman
Lynn Welch
Storm Whaley
James M. Witt
Sylvan H. Wittwer
Robert Wood
Robert R. Wright
Nanci Wrubel
Fred Weingarten
Karl Yordy
James Zwolenik

☆U.S. GOVERNMENT PRINTING OFFICE: 1982–555-119

National Science Foundation
Washington, D.C. 20550

Official Business

PENALTY FOR PRIVATE USE $300

RETURN THIS COVER SHEET TO ROOM 233, IF YOU DO NOT WISH TO RECEIVE THIS MATERIAL ☐, OR IF CHANGE OF ADDRESS IS NEEDED ☐ (INDICATE CHANGE, INCLUDING ZIP CODE).

Postage and Fees Paid
National Science Foundation

U.S.MAIL

SPECIAL FOURTH CLASS RATE
BOOK

NSB 80-1