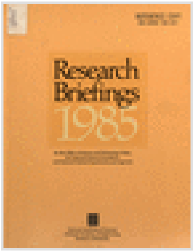
FROM THE ARCHIVES



Research Briefings, 1985 (1985)

Pages
112

Size
8.5 x 11

ISBN
0309035856

Committee on Science, Engineering, and Public Policy; National Academy of Sciences; National Academy of Engineering; Institute of Medicine

Research Briefings 1985

Research Briefings 1985

for the Office of Science and Technology Policy,
the National Science Foundation,
and Selected Federal Departments and Agencies

Committee on Science, Engineering,
and Public Policy
National Academy of Sciences
National Academy of Engineering
Institute of Medicine

Order from
National Technical
Information Service,
Springfield, Va.
22161
Order No. PB86148582

NATIONAL ACADEMY PRESS
Washington, D.C. 1985

National Academy Press 2101 Constitution Avenue, NW Washington, DC 20418

NOTICE: The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the furtherance of science and technology for the general welfare. The terms of its charter require the National Academy of Sciences to advise the federal government upon request within its fields of competence. Under this corporate charter, the National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively.

The Committee on Science, Engineering, and Public Policy is a joint committee of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. It includes members of the councils of all three bodies.

This work was supported by the National Science Foundation under Contract PSP791304.

Library of Congress Catalog Card Number 85-61729

International Standard Book Number 0-309-03585-6

Printed in the United States of America

Committee on Science, Engineering, and Public Policy

LEON T. SILVER, William M. Keck Foundation Professor of Geology, Division of Geological and Planetary Sciences, California Institute of Technology (*Chairman*)
*LINDA H. AIKEN, Vice President for Research, The Robert Wood Johnson Foundation, Princeton, New Jersey
RICHARD C. ATKINSON, Chancellor, University of California, San Diego
JACOB BIGELEISEN, Leading Professor, Department of Chemistry, State University of New York, Stony Brook
FLOYD E. BLOOM, Director and Member, Division of Pre-Clinical Neuroscience and Endocrinology, Scripps Clinic and Research Foundation, La Jolla, California
W. DALE COMPTON, Vice President, Research, Ford Motor Company, Dearborn, Michigan
EMILIO Q. DADDARIO, Wilkes, Artis, Hedrick and Lane, Attorneys at Law, Washington, D.C.
GERALD P. DINNEEN, Vice President, Science and Technology, Honeywell, Inc., Minneapolis, Minnesota
ALEXANDER FLAX, President Emeritus, Institute for Defense Analyses, Alexandria, Virginia
ZVI GRILICHES, Professor, Department of Economics, Harvard University, Cambridge, Massachusetts
PHILIP LEDER, John Emory Andrus Professor and Chairman, Harvard Medical School, Cambridge, Massachusetts
GARDNER LINDZEY, President and Director, Center for Advanced Study in the Behavioral Sciences, Palo Alto, California
EDWARD A. MASON, Vice President, Research, Standard Oil Company (Indiana), Naperville, Illinois
DANIEL NATHANS, Professor, Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland
GILBERT S. OMENN, Dean, School of Public Health and Community Medicine, University of Washington, Seattle
HERBERT A. SIMON, Richard King Mellon University Professor of Computer Science and Psychology, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania

*Term expired December 31, 1984

Preface

In 1982 the Office of Science and Technology Policy (OSTP) asked the National Academy complex[1] to develop a series of briefings aimed at identifying research areas of unusual scientific opportunity within specific fields of science. The first set of briefings covered seven fields, and were presented in both oral and written form to senior officials of OSTP and the National Science Foundation (NSF) in late 1982. The briefings were presented subsequently to senior officials of other interested federal departments and agencies, and published as a collection by the National Academy Press. Responsibility for developing the briefings was given to the Committee on Science, Engineering, and Public Policy (COSEPUP), a joint unit of the National Academies of Sciences and Engineering and the Institute of Medicine.

Thus began the program of annual research briefings that with the publication of this volume will have completed four rounds and covered 28 topics. Topic names and definitions for 1985 are listed below, along with a listing of topics in previous years.

[1]Together, the National Academies of Sciences and Engineering and the Institute of Medicine are known as the National Academy complex.

1985[2]

1. *Pain and Pain Management*: Research directed toward enhanced understanding of the basic biological and behavioral phenomena, and their interactions, that underlie pain and analgesia and that lend insight for improved clinical management of pain.
2. *Biotechnology in Agriculture*: Research directed toward increasing the productivity and efficiency of American agriculture using the techniques of modern biotechnology.
3. *Computer Vision and Pattern Recognition*: Mathematical and other research directed toward enhanced capabilities for recognition of objects and geographical locations through comparison of sensor-generated data and stored image information.
4. *Weather Prediction Technologies*: Examination and evaluation of the impact on short- and long-term weather prediction of new

[2]Presentation of briefings on seven topics in 1985 does not imply that other topics not presented are considered to be less important to the progress of U.S. science and technology. No priority setting among fields is implied. The list of 1985 topics reflects the fact that only a limited number of timely topics can be covered in a given year.

technological developments for collection and processing (including mathematical modeling) of meteorological data.

5. *Remote Sensing of the Earth*: Research directed toward enhanced capabilities in all spectral regions for study of the earth from satellite-borne instruments, including observation of land and water surfaces and the atmosphere.

6. *Ceramics and Ceramic Composites*: Research directed toward enhanced understanding of the relationships between microscopic behavior and macroscopic properties of ceramics and ceramic composites, taking into account structure composition and processing history plus environmental conditions to which these materials are exposed. The ultimate goal is to understand these relationships such that ceramic materials can be produced with controlled reproducible properties and their performance under in-service conditions predicted.

7. *Scientific Frontiers and the Superconducting Super Collider*: Examination and evaluation of the scientific opportunities presented by the availability of the proposed superconducting super collider, including opportunities in elementary particle physics and impact on other frontier fields of science.

1984[3]

1. Computer Architecture
2. Information Technology in Precollege Education
3. Chemical and Process Engineering for Biotechnology
4. High-Performance Polymer Composites
5. Biology of Oncogenes
6. Interactions Between Blood and Blood Vessels (Including the Biology of Atherosclerosis)
7. Biology of Parasitism
8. Solar-Terrestrial Plasma Physics
9. Selected Opportunities in Physics

1983[4]

1. Selected Opportunities in Chemistry
2. Cognitive Science and Artificial Intelligence
3. Immunology
4. Solid Earth Sciences
5. Computers in Design and Manufacturing

1982[5]

1. Mathematics
2. Atmospheric Sciences
3. Astronomy and Astrophysics
4. Agricultural Research
5. Neuroscience
6. Materials Science
7. Human Health Effects of Hazardous Chemical Exposures

Topics are generally selected by Dr. George Keyworth, Science Advisor to the President, after consultation with representatives of NSF and COSEPUP. The ceramics topic in 1985 was specifically identified by the NSF as one of timely interest. Each briefing is developed by a panel of experts charged with assessing the status of their field and identifying those research areas within the field likely to return the highest scientific dividends as a result of additional near-term federal investment. Prior to oral delivery of the briefings and their publication they are carefully reviewed by COSEPUP. Financial support for all four rounds of briefings has been provided by NSF.

In its brief history the research briefing activity has become an important channel for

[3]Published as *Research Briefings 1984*, Washington, D.C.: National Academy Press, 1984.

[4]Published as *Research Briefings 1983*, Washington, D.C.: National Academy Press, 1983.

[5]Published as *Research Briefings*, Washington, D.C.: National Academy Press, 1983.

communication between the U.S. research community and those responsible for federal funding of research and development. In the words of Dr. Keyworth,[6] "The impacts . . . have been in reinforcing perceptions, in strengthening resolve, and in clarifying the often confusing multiplicity of inputs . . ." converging on the federal agencies. Dr. Keyworth has also pointed out that one of the briefings[7] ". . . led almost directly and quickly to an important new program of Engineering Research Centers in the National Science Foundation." Overall, it is his perception that ". . . the briefings have become increasingly helpful in formulating science policy."

Needless to say, the briefings could not have been prepared without the voluntary contributions of the hundreds of scientists and engineers who gave generously of their time. COSEPUP is extremely grateful to those who have served on the briefing panels and helped to develop an important new means of informing government officials.

Leon T. Silver, *Chairman*
Committee on Science, Engineering, and Public Policy

[6] *New Pathways in Science and Technology. Collected Research Briefings 1982-84*, New York: Random House, Inc., 1985.
[7] "Computers in Design and Manufacturing," 1983.

Report of the Research Briefing Panel on Remote Sensing of the Earth

Research Briefing Panel on Remote Sensing of the Earth

D. James Baker (*Chairman*), Joint Oceanographic Institutions, Inc., Washington, D.C.
Francis Bretherton, National Center for Atmospheric Research, Boulder, Colo.
Kevin Burke, Lunar and Planetary Institute, Houston, Tex.
Charles L. Drake, Dartmouth College, Hanover, N.H.
Richard Goody, Harvard University, Cambridge, Mass.
H. William Menard, Scripps Institution of Oceanography, La Jolla, Calif.
Berrien Moore III, University of New Hampshire, Durham
Ronald Prinn, Massachusetts Institute of Technology, Cambridge
John Steele, Woods Hole Oceanographic Institution, Woods Hole, Mass.
Verner E. Suomi, University of Wisconsin, Madison
Wilford F. Weeks, Cold Regions Research and Engineering Laboratory, Hanover, N.H.
Carl Wunsch, Massachusetts Institute of Technology, Cambridge

Staff

Dean P. Kastel, *Staff Director*, Space Science Board
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

Richard Johnson, *Liaison Representative*, Office of Science and Technology Policy

Report of the Research Briefing Panel on Remote Sensing of the Earth

SUMMARY OF RECOMMENDATIONS

Until the flight of satellites, there were no techniques for long-term, global, synoptic measurements of processes in the atmosphere, oceans, and solid earth. Now we are on the verge of establishing a system of remote sensing instruments and earth-based calibration and validation programs that could provide such a data set. With the concurrent development of numerical models that can run on supercomputers, we have the potential of achieving significant advances in understanding the state of the earth, its changes, feedbacks, interactions, and global trends on time scales ranging from days to centuries. To reach this potential, this panel recommends that:

For new missions and new instruments:

High priority be given to the TOPEX mission and the Ocean Color Imager for new starts in FY 1987 and to GRM for a new start in FY 1988, so that simultaneous observations of major processes on the earth will be available by the early 1990s;

Immediate and adequate funding be given to the development of techniques based on satellite remote sensing for the accurate and global measurement of precipitation;

Funds be made available for the development and flight of improved surface imaging spectrometers and tropospheric chemical sensors;

Space be made available on operational missions and Explorer-type earth science missions for testing of new techniques and instruments, and continued and adequate support be provided for the development of an expanded space-based earth observation effort for the 1990s and beyond.

For operational satellite measurements:

The United States as a matter of high priority maintain and continue to upgrade the existing system of civilian operational weather measurements from two polar-orbiting and two geostationary satellites;

Intercalibration of instruments to be used on successive missions that could yield long-term records be an integral part of operational mission planning.

For access of data:

The establishment and maintenance on a continuing basis of adequate, long-term global data bases and correlative observations on specific environmental parameters required to assess both natural variability and human impact on the environment be recognized as a high national priority;

NOAA, NASA, and other appropriate agencies be funded now to adequately develop their data bases and to make them accessible to researchers and other users (This includes archiving the existing satellite data from both U.S. and non-U.S. missions in a form convenient for outside users, and giving proper attention to documenting, storing, reducing, and distributing these data sets);

High priority be given to identifying those data sets of particular importance to long-term changes on the earth, for example, early LANDSAT data, and providing funds to preserve and make available these data to researchers and other users;

Data management be addressed as an integral part of any space mission, and that the views of the interested scientific community be sought in this process to establish requirements for handling, processing, and storing the significant quantities of data expected from future U.S. and non-U.S. missions.

For required complementary measurements and analyses:

The panel emphasizes the importance of a complementary program of earth-based calibration and validation studies, and the critical need for sustaining an overall and coordinated program of space- and earth-based studies together with computer-based modeling. If we are to achieve our goal of understanding the earth system, neither space- nor earth-based measurements can stand alone; the two must be developed simultaneously and tied together with models.

For agency roles and development of a broader community:

The space-based earth science programs of the involved agencies be jointly reviewed and coordinated at the interagency level;

The National Research Council be considered as a channel to provide advice on such a program through the appropriate board and committees;

Basic research funds and facilities be made available as required on a peer-reviewed basis;

Faculty development awards and institutional graduate student support be made available for the development of an adequately trained community.

INTRODUCTION—REMOTE SENSING OF GLOBAL CHANGE

Twenty-five years ago the new meteorological satellites gave atmospheric scientists their first global view of atmospheric processes. Thirteen years ago the first imaging instruments for land processes were flown, and just seven years ago the first ocean satellite provided three months of information on ocean processes and the shape of the geoid. Today satellite remote sensing technology for the earth extends to a wide band of electromagnetic radiation ranging in wavelengths from the microwave to the ultraviolet (see Figure 1).

The emerging global view of the earth and its processes has led us to the verge of dramatic new understanding of the state of planet Earth and its changes on time scales varying from hours and days to months, years, centuries, and millennia. With this new understanding, and with a growing data base on earth processes, we will be able to address the critical questions of predictability and to identify the feedbacks and interactions in the system that give rise to global change. At the same time we will be able to establish the natural variability of the system that is the context for interpretation of developing trends.

Short-term predictability has been treated in other reviews from the Committee on Science, Engineering, and Public Policy: understanding the processes that cause earthquakes by the Research Briefing Panel on Solid Earth Sciences in 1983, and under-

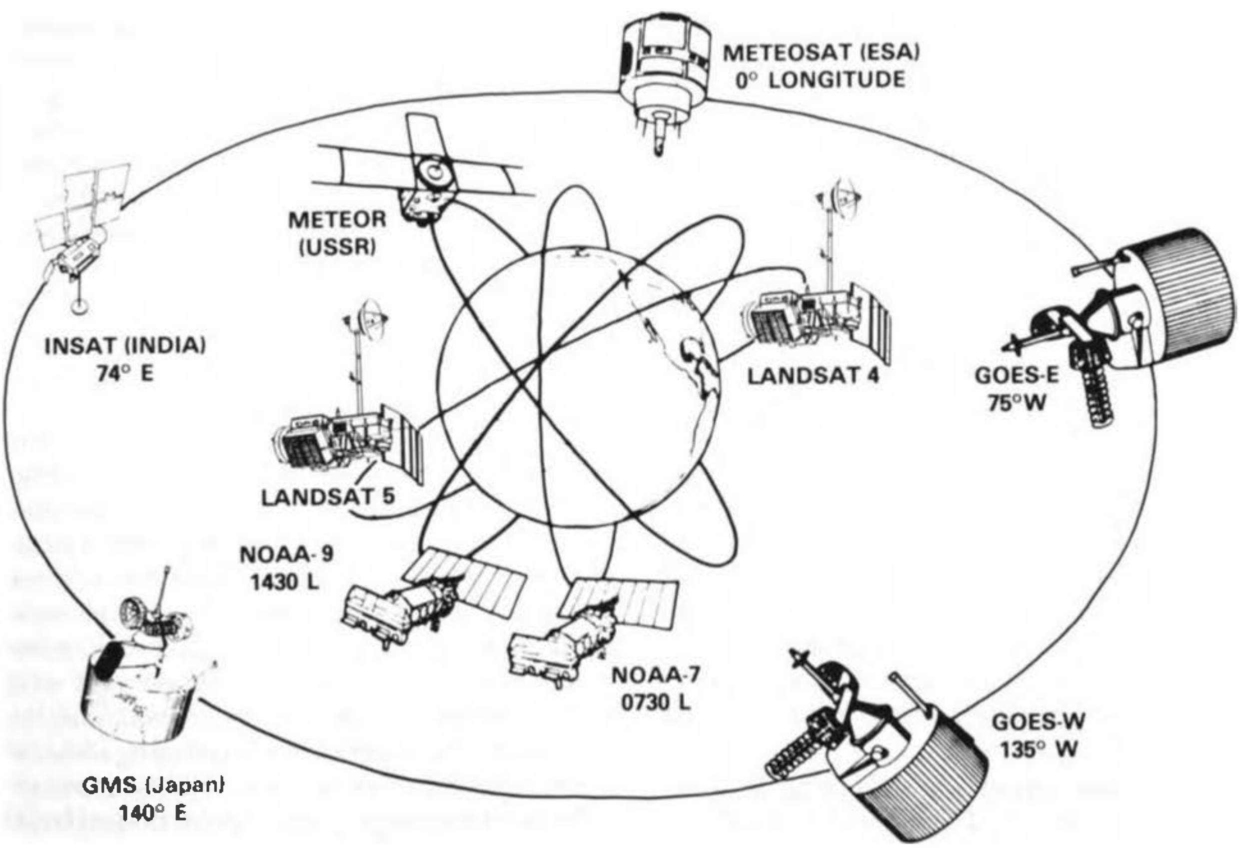


FIGURE 1 Operational Civil Earth Observation Satellites: Polar-orbiting and geostationary satellites monitor weather and surface conditions over the entire globe. SOURCE: NOAA, National Environmental Satellite, Data, and Information Service.

standing the physics of weather prediction on time scales of hours to weeks by the Research Briefing Panel on Weather Prediction Technologies this year. This report identifies special opportunities for advances in understanding global environmental change in time periods ranging from days to centuries. Global change, once it becomes obvious, is likely to be irreversible on human time scales. The ability to diagnose what is happening at the earliest possible stage and to perceive the synergism between seemingly unrelated events may be critical in helping the peoples of the world adjust to changing circumstances and to far-reaching environmental constraints.

The magnitude of global change has only recently become apparent through new remote sensing technology. For example, we now know that the tropical oceans interact with the global atmosphere to cause the El Niño climate anomaly. Ejections of material from volcanoes can lead to global cooling, whereas the increased concentration of gases such as carbon dioxide, methane, and chlorofluoromethanes in the atmosphere can lead to global temperature change. Long-term decreases in upper atmospheric ozone caused by increases in oxides of nitrogen and chlorine can cause increased penetration of ultraviolet radiation, leading to greater incidence of skin cancer, also a global effect. The loss of forests and desertification can be destabilizing, causing climatic anomalies, soil erosion, and consequent loss of agricultural productivity. The cycles of life-supporting elements—carbon, sulfur, phosphorus, nitrogen, and water—are all affected by storage and transfer between the atmosphere, the ocean, the biota, and the solid earth.

To advance our understanding of the causes and effects of global change, we need new observations of the earth. These measurements must be global and synoptic, they must be long-term, and different processes such as atmospheric winds, ocean currents, and biological productivity must be measured simultaneously. We have learned that major advances in earth sciences have come from syntheses of new ideas drawn from such global synoptic observations. The synthesis of plate tectonics from large-scale data is a major step in understanding how the solid earth works; the understanding of the dynamics of large-scale circulation of the atmosphere that comes from global observations has permitted a significant increase in the accuracy of weather prediction. Now we must take the next steps.

Long-term continuity is also crucial. A 20-year time series of the crucial variables would provide a significant improvement in our understanding. Twenty years cover two solar sunspot cycles; it is the period over which we can expect the temperature change due to radiatively active gases to be larger than the natural system noise; it encompasses the eruptions of 5 to 10 volcanoes and the occurrence of 2 to 5 El Niños; and it is the period over which we can expect to see the major effects of deforestation. Finally, we note the need for simultaneity. If we are to make progress in understanding the earth as a system it is essential that we make physical, chemical, and biological observations all at the same time since the physics, chemistry, and biology are all interrelated.

Until the advent of satellites, we had no techniques that could satisfy the needs for long-term, global, synoptic measurements of different processes on the earth. Now we are on the verge of establishing a global system of remote sensing instruments and earth-based calibration and validation programs. Together, these space- and earth-based measurements can provide the necessary data. With the concurrent development of numerical models that can run on supercomputers, we have the potential of achieving significant advances in understanding the state of the earth, its changes, feedbacks, interactions, and global trends on time scales of years to centuries.

NEW MISSIONS AND NEW INSTRUMENTS

It is crucial that data records contain the proper parameters as well as provide long-term continuity. Thus, in addition to the continuance and improvement of the operational weather programs (see below), we need a continuing series of research missions to test new instruments and to measure processes that cannot be defined with data from the limited operational satellite system. The support of a continuing series of research missions is essential to the development of an adequate data set for understanding the earth. The long intervals between launches of research satellites in the earth sciences in the past has not allowed a proper follow-up of earlier developments and has broken the chain of continuous measurements.

With the heritage of the past decade, we now have flight-proven instruments for making global measurements never before undertaken and we can identify areas of development that will make a major change in our understanding of the state of the earth. It is essential that there be a long-term national commitment to satellite missions that will carry the instruments that are now ready, and we must now begin to develop the necessary new instruments and techniques.

New missions identified as high priority are the Ocean Topography Experiment (TOPEX) and the Geopotential Research Mission (GRM). An important new instrument ready for flight is the Ocean Color Imager. Areas of development include spectral imaging instruments for land biology and geology and for tropospheric gases, particularly carbon monoxide, and new techniques for measurement of precipitation. In addition, the panel sees important opportunities for Explorer-type missions for earth science; two of particular interest are the measurements of the large-scale magnetic field and the test of tropical precipitation and tropospheric carbon monoxide measurement techniques.

The Ocean Topography Experiment (TOPEX)

The short-lived SEASAT mission showed the oceanographic community that a technological revolution in ocean observation was finally possible. For more than 100 years, oceanography has had to rely upon ships and isolated, short-duration instruments to gain information about a global, turbulent fluid. Until the advent of the sensors on SEASAT, the density of observations possible over and within the ocean necessarily failed by several orders of magnitude to meet even the most rudimentary of the known sampling requirements of a turbulent fluid, one where all scales of motion were known to be important.

SEASAT showed that modern microwave electronics could provide global, continuous measurements of surface atmospheric wind stress, the most important field of forcing of the ocean, and of the dynamical response of the ocean through the movements of its upper surface. Plans are now in place for the Navy Remote Ocean Sensing System (NROSS) and the European Space Agency's ERS-1 to provide the wind stress measurements beginning in 1990; the precision ocean topography measurements needed for ocean currents will be provided by the proposed NASA mission TOPEX.

The U.S. and world oceanographic communities have spent the past five years constructing an orderly approach to getting the new observational technologies into place. Much of the planning has taken place in the context of the World Climate Research Program, which includes the Tropical Ocean-Global Atmosphere Program (TOGA) and the World Ocean Circulation Experiment (WOCE). WOCE planning, which includes global satellite and *in situ* programs of ocean measurements, has led to an orderly sequence for the proposed launch of the requisite spacecraft (NROSS, TOPEX [in its most desirable form, as the U.S./French mission TOPEX/POSEIDON], GRM [see below], and ERS-1). The flight of these spacecraft, phased approximately over the period 1990 to 1996, would be accompanied by the international application of all the modern physical and chemical oceanographic techniques on a global scale.

Figure 2 shows the dramatic improvement in the estimate of the meridional heat flux in the ocean anticipated with the application of the combination of TOPEX altimetry and WOCE tropical carbon-14 tracer data.

The Geopotential Research Mission (GRM)

The Geopotential Research Mission is designed to measure variations in the gravity and magnetic field over the entire earth to a resolution of 2 milligals and 2 nanotesla for 100-km half wavelengths. Because it will provide a much improved geoid, to which sea surface heights are referred, the GRM is needed to allow full realization of the results of the contemporary TOPEX mission. GRM is proposed to fly during the NROSS/TOPEX flight period.

Gravity measurements from GRM are expected to reveal, for the first time, the pattern of underlying mantle circulation. This information will show how the smaller scales of mantle convection interact with the large-scale plate structure of the lithosphere. Tantalizing patterns of mantle circulation beneath the oceans have been discerned from SEASAT altimetry, and seismic tomography is beginning to reveal the deeper structure of the mantle. As the global digital seismic network (GSN) becomes operational, the data

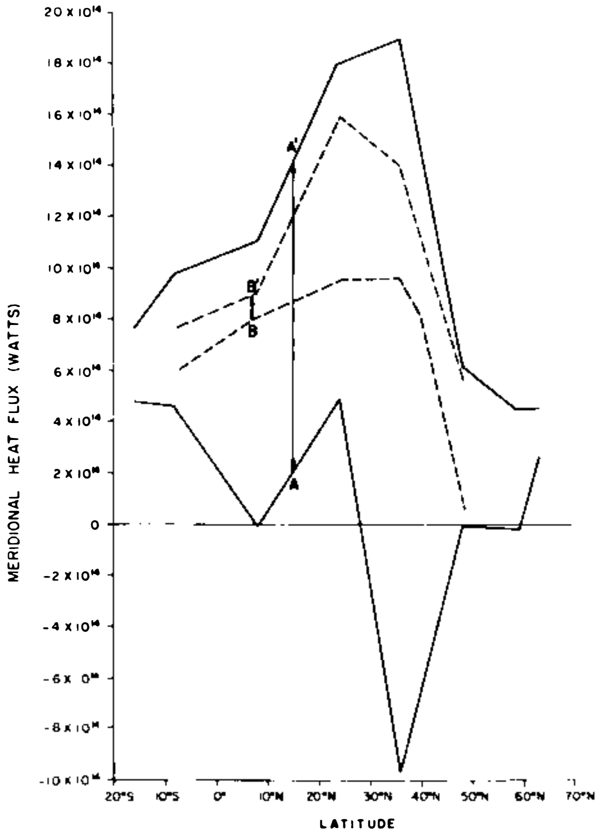


FIGURE 2 Expected improvement of estimate of ocean heat flux with new data sets: A-A′ is the estimated uncertainty for the annual average meridional heat flux in the North Atlantic using existing data; B-B′ is the uncertainty expected with the application of satellite altimetry and *in situ* carbon-14 tracer data. SOURCE: C. Wunsch, Massachusetts Institute of Technology.

sets from GRM, SEASAT, and GSN can be combined to reveal for the first time the relation between plate motion and the underlying mantle circulation.

Gravity fields measured from GRM will uniquely illuminate the mass distribution under the Himalayas, Tibet, and neighboring areas of Asia where the collision of India with Eurasia is currently being accommodated by the elevation of the world's highest mountains and by the extrusion of China toward the Pacific Ocean. Collisions of this kind have been a major element in the evolution of the continents over the last four billion years; acquisition of the GRM gravity field is essential to full description and understanding of this complicated process. A complementary data set will be acquired by GRM over the high Andean chain where calculated mass distributions can be used to establish how active subduction and volcanism operate.

The 160-km altitude orbit, essential for the GRM gravity measurements, also provides an unprecedented opportunity to fly a magnetometer in very low earth orbit. This instrument will not only measure the large-scale magnetic field, but will also establish the distribution of anomalies in the field at the 100-km scale that are attributable to ferromagnetic minerals in the continental crust. Over much of the continents this information is otherwise unattainable; thus acquisition of the GRM magnetic field represents an essential element in establishing earth history. A magnetic field explorer mission, at higher altitudes, will measure the large-scale magnetic field for studies of its generation and changes.

The Ocean Color Imager

A major concern for earth science in the next decade and beyond is an understanding of biogeochemical cycles; at decadal time scales the cycles play a major part in critical processes such as the carbon, nitrogen, and sulphur exchanges. The oceans' role in these cycles is known qualitatively and operates in part through vertical exchanges: there is strong and rapid coupling between surface and deep water through particle fluxes that carry organic matter. Large-scale connections were recently seen in the changes in atmospheric carbon dioxide occurring at the time of the 1982–1983 El Niño, which produced massive changes in physical, chemical, and biological properties in the Pacific.

Investigations of these vertical fluxes are under way through studies of chemical distributions in the ocean; by direct measurements of particle transports and the rates of transformation between particulate and soluble phases; and through modeling of the combined physical, chemical, and biological dynamics. These studies depend on a knowledge of the primary production of organic matter in the upper layer of the ocean. We have methods to determine these rates by shipboard experiments at discrete locations, but we require a global perspective of this process and its variability in space and time.

Satellite measurements of ocean color can provide this synoptic view. The Coastal Zone Color Scanner (CZCS) was intended for regional (and fisheries) studies, but has also proved capable of use for open ocean study. Algorithms have been developed to relate the remotely sensed data to chlorophyll, the primary pigment for photosynthesis. In turn, these estimates can be correlated with measurements of primary production. The proposed Ocean Color Imager (OCI), with more spectral channels, will provide more accurate measurements of chlorophyll. The new data will yield a better understanding of the essential links between physical changes in the ocean, biological production (including fisheries), and chemical fluxes to deep water. The integration of these processes using satellite data is essential for furthering our understanding of general biogeochemical cycles. It will be possible to test hypotheses about the role of these cycles, not only in their coupling with ocean dynamics but also in controlling the overall planetary system, including feedback to atmospheric and terrestrial components.

These concepts and the necessary technologies are now being formulated into a plan for a Global Ocean Flux Study, which can provide the links between physical oceanography programs and the corresponding long-term atmospheric and terrestrial programs. The studies on basic production are relevant to a further initiative on recruitment processes, and the OCI is an essential and central element in these overall plans.

Measurement of Precipitation and Evaporation

Accurate measurements of the rates of precipitation, evaporation, and evapotranspiration over the global land and ocean surfaces are essential for understanding the coupled earth-land-ocean-atmosphere system. Evaporation is closely tied to temperature, wind stress, and humidity. Strong efforts are being made to improve these measurements from remote instruments. Unfortunately, precipitation is highly variable and poorly measured especially over the ocean; we lack a reliable technique for this measurement.

There are two problems in obtaining the global coverage of rainfall desired for climatic studies: sampling and measurement. The traditional source of rain data, a rain gauge network, is unavailable over the oceans. Because of the convective nature of the rainfall, island stations are not representative of the surrounding ocean, and ship station rainfall observations are sparse as well as notoriously unrepresentative and unreliable.

During the Atlantic Tropical Experiment of the Global Atmospheric Research Program it was shown that satellite microwave data were highly correlated with calibrated ship radar when both were observing the same area. Unfortunately the satellite systems were turned on only every other orbit and had a swath width less than the interorbit gap. Thus, while microwave multifrequency radiometers on polar-orbiting satellites appear to meet the measurement accuracy criteria, they fail on adequate sampling.

Geostationary satellites sample much of the earth continuously and the tropical belt completely. They provide information on where it is raining, but not on how much. Several rain estimation schemes have been devised; for example, there is good evidence that the cloud height as measured by geostationary satellites is closely related to rain rate for convective rain. Unfortunately, the errors in this technique are too large.

There are a number of promising techniques that need support for development now. Such techniques follow two schemes, scattering or absorption. The scattering scheme, useful at the higher microwave frequencies, is largely empirical. On the other hand, the theory of microwave attenuation due to rainfall is in much better order. Here the key uncertainties remaining are the thickness of the rain layer and the presence or absence of the ice phase. A satellite equipped with a suitable radar can remove this uncertainty and greatly improve the accuracy of microwave observations. Adequate sampling from a polar orbiting satellite would still be a problem.

In summary, the requirements for accurate rainfall measurement are demanding, but we believe that satellite platforms now in orbit could be equipped to test and improve our measurement capability. The development and flight of precipitation measurement techniques on operational satellites and on Explorer-type missions is of high priority for any earth sciences program.

Measurements With Improved Imaging Spectrometers

The most broadly useful instrument for geology and biology among the new capabilities envisioned for Earth observation in the 1990s is a moderate-resolution imaging spectrometer (MODIS). The spatial resolution would be similar to that presently available (about 1 km), but the focus would be on greatly increasing the spectral information. We now know that important information about the earth and its biology can be found by observing precisely and simultaneously in many parts of the visible and infrared spectrum. For example, vegetative stress, including chemical imbalances induced from acid rain, can be detected as can the subtle composition of minerals through the spectra of hydrous compounds.

The MODIS instrument would also offer a major enhancement to ocean color measurements begun by the Coastal Zone Color Scanner and to be extended by the Ocean Color Imager, and it would greatly expand the capabilities offered by the current operational radiometers. The new technique is now available and ready for development for use in space. As part of this development there is a need for experimental use of the spectral techniques over test areas.

These techniques promise a new dimension for remote sensing from space for geologic structure and geomorphology that has been successful since the time of the first LANDSAT mission (1972). Continuing improvements in technique (e.g., stereo images, nonnadir look angles, higher spectral resolution, imaging at a variety of times of day) are likely to lead to increasing demand for the use of these tools and also of related spaceborne photography (e.g., large-format camera products).

Tropospheric Chemical Sensors

We now know that the chemistry of the troposphere is controlled to a considerable extent by biological and human processes and hence is not static but changing. Satellites have played important roles in the study of upper atmosphere chemistry and will continue this role with the launch of the Upper Atmosphere Research Satellite (UARS) in 1988. In contrast, satellites have not yet played a major role in tropospheric chemistry because of difficulties in remote sensing of many tropospheric species. However, carbon monoxide, which is one of the most important species controlling the oxidation state of the global atmosphere, is detectable from space. We urgently need to know the global distribution and trend of carbon monoxide, but this gas is so variable in space and time that satellites appear to provide the only practical way of meeting this need. An exploratory sensor has been flown on the space shuttle, which shows the feasibility of CO measurements from space. A more sophisticated sensor capable of measuring the vertical profile of CO over the globe and flown on a polar orbiter would provide data of crucial importance to our understanding of global tropospheric chemistry.

The panel notes the importance both of the new research missions and of the development of new techniques and instruments to build a base for an expanded space-based earth observation effort for the 1990s and beyond. Specifically, the panel recommends that:

- *High priority be given to the TOPEX mission and the Ocean Color Imager for new starts in FY 1987 and to GRM for a new start in FY 1988, so that simultaneous observations of major processes on the earth will be available by the early 1990s.*
- *Immediate and adequate funding be given to the development of techniques based on satellite remote sensing for the accurate and global measurement of precipitation.*
- *Funds be made available for the development and flight of improved surface imaging spectrometers and tropospheric chemical sensors.*
- *Space be made available on operational missions and Explorer-type earth science missions for test of new techniques and instruments, and continued and adequate support be provided for the development of an expanded space-based earth observation effort for the 1990s and beyond.*

OPERATIONAL WEATHER AND ICE MEASUREMENTS—THE RESEARCH BASE

The U.S. civil operational earth-observing satellite systems, together with their earth-based calibration and validation programs, provide a global data base for meteorology, oceanography, solid-earth geophysics, and solar-terrestrial science. From these sources, federal agencies provide operational services critical to the protection of life and property, the national economy, energy development and distribution, and global food supplies.

At the same time, this data base is fundamental to research on the state of the earth and its predictability. Thus, it is essential that the operational satellite measurement system be continued and improved to provide homogeneous, accurate, and timely data. The panel notes that its recommendations, as well as those from the Panel on Weather Prediction Technologies, are based on the assumption that the operational systems now in place will continue.

Today, two U.S. polar-orbiting weather satellites and two U.S. geostationary satellites monitor weather and surface conditions all over the globe. Two U.S. earth resources satellites (LANDSATs) are also in polar orbit. These U.S. systems are complemented by a number of foreign Earth-observing satellites, and more foreign satellites will be launched in the latter half of the 1980s.

The polar-orbiting weather satellites provide operational coverage of the entire earth four times per day, measuring temperature and humidity in the Earth's atmosphere, surface temperature, cloud cover, water-ice boundaries, and proton and electron flux near the earth. They receive, process, and distribute data from stations distributed globally. Derived polar satellite products include information on ice, vegetation, snow, and tropical cyclones.

Improvements needed for future polar-orbiting satellites include additional channels in the sounding units and added capability in the radiometers for better atmospheric moisture measurements, better estimation of worldwide vegetation levels, and operational measurement of solar and terrestrial radiation levels.

The geostationary satellites provide synoptic hemispheric coverage that extends well into the southern hemisphere. Their sounding/imaging instruments make day and night observations of weather in the coverage area and relay data from surface collection points to processing centers. The geostationary sounders provide a complementary measurement to the polar-orbiters with their synoptic vantage point, and also monitor the magnetic field of the earth, the energetic particle flux, and gamma ray emissions from the sun.

Improvements in the next generation of geostationary satellites include replacement of the current sounding/imaging instrument with separate instruments. A stabilized spacecraft would increase the available scanning time, and detector arrays rather than single-point detectors could scan swaths of the earth and so reduce the time for complete area coverage while retaining the same size field of view and multichannel spectral data input.

Lack of conventional observational stations in the polar regions still represents a major data gap. The synoptic observations required are only possible with satellite remote sensing. Present satellite systems have already contributed significantly to improved observations even though sensors in the visible and infrared are limited by clouds and darkness. The most useful remote sensing system for sea ice studies currently deployed is a passive microwave imager which provides all-weather, low-resolution imagery. However, an adequate validation of the algorithm, or transfer function, used to calculate ice conditions from this data is still lacking.

The next step in ice-observing capabilities will come with the launching of an improved passive microwave system on a U.S. Defense Meteorological Satellite in 1986. Then we look to the flight of synthetic aperture radar systems on European, Canadian, and Japanese satellites during 1989 to 1995. For the United States to receive maximum benefit from the radar systems, it will be necessary to establish ground truth observations during shuttle test missions in 1987 and 1988 to validate the interpretation of this imagery over ice-covered oceans. In addition, the schedule for establishment of receiving stations for the non-U.S. satellites must be maintained. The high resolution of synthetic aperture radar (tens of meters) makes it valuable also for open ocean and land surface studies.

The need for all these operational measurements has been firmly emphasized internationally. At the seventh meeting of the Economic Summit of Industrialized Nations the heads of state established a Working Group on Technology, Growth, and Employment; remote sensing from space is the topic of a panel chaired by the United States. The summit panel has already emphasized the importance of long-term continuity of operational meteorological satellite data.

In view of the fundamental importance of the data that is collected by the operational weather satellites to the research community, the panel recommends:

- *That the United States as a matter of high priority maintain and continue to upgrade the existing system of civilian operational weather measurements from two polar-orbiting and two geostationary satellites.*

In addition, to ensure that long-term data records are as homogeneous as possible, the panel recommends that:

- *Intercalibration of instruments to be used on successive missions that could yield long-term records be an integral part of operational mission planning.*

ACCESSING THE DATA

Over the next 10 to 15 years we expect to see a major increase in data volume from the operational remote sensing systems and the new research satellite programs, coming largely from requirements for higher spatial and spectral resolution, repetitive measurements, long-term data sets, and global processes. At the same time, we find now that access to existing remotely sensed data is not adequate for research purposes. Thus a major high-leverage investment could be made now in the improvement of the data management, archiving, and distribution of operational satellite data for research users.

The effective management of large data bases is a general problem in space science, and has been addressed in a number of reports. In essence, it has proved almost impossible for the researcher to "drink from the firehose" of satellite data streams. It will be far more difficult to digest data from a number of different sources. In the panel's view, the scientific exploitation of Earth observations are likely to be limited as much by the inability to manipulate and distribute efficiently large and diverse data sets as by sensor technology. The challenge lies in information development and distribution; a challenge that can be met with new high-technology computing and communication techniques if adequate funds are made available. The several planning efforts now under way for high-speed data networks to link supercomputers, data centers, and university research laboratories need continued support.

The panel's recommendations for data management are in two parts: one on current data, and one on data that will come from satellites that will fly in the next decade. In terms of current data, the panel recommends that:

- *The establishment and maintenance on a continuing basis of adequate, long-term global data bases and correlative observations on specific environmental parameters that are required to assess both natural variability and human impact on the environment be recognized as a high national priority;*
- *NOAA, NASA, and other appropriate agencies be funded now to develop their data bases adequately and to make them accessible to researchers and other users (This includes archiving the existing satellite data [from both U.S. and non-U.S. missions] in a form convenient for outside users, and giving proper attention to documenting, storing, reducing, and distributing these data sets);*
- *High priority be given to identifying those data sets of particular importance to long-term changes on the earth, for example, early LANDSAT data, and providing funds to preserve and make available these data to researchers and other users.*

Looking ahead, the panel supports the recommendations of the Space Science Board's Committees on Earth Sciences and on Data Management and Computation that:

- *Data management be addressed as an integral part of any space mission, and that the views of the interested scientific community be sought in this process to establish requirements for handling, processing, and storing the significant quantities of data expected from future U.S. and non-U.S. missions.*

Specific developments that need funding in the next two years include high-speed parallel processors and other large computational machines that will be required for the large data sets to come, and the development of communications networks linking data centers, supercomputers, and university research laboratories.

REQUIRED COMPLEMENTARY MEASUREMENTS AND ANALYSES

The measurements obtained from satellites must be complemented by measurements in the field to reliably calibrate, validate, and interpret the satellite data. In addition, many of the important scientific problems that can be addressed using satellite measurements also require measurements taken on the surface of the earth or from aircraft. It is essential that the program for remote sensing of the earth from space include such complementary field measurements.

In the atmosphere, measurements of basic meteorological variables (temperature, pressure, humidity) from balloons have long been recognized as essential to the accurate interpretation of satellite data. In addition, while satellite data are useful for the study of storms, ground-based and airborne measurements provide data on variables such as wind velocities at levels of accuracy that cannot be attained from satellites.

Also in the atmosphere, many chemically important trace species are not measurable from space with current techniques. However, the concentrations of these species can be obtained through current ground-based or airborne techniques. Such information is essential to understanding the chemistry of species such as stratospheric ozone and tropospheric carbon monoxide, which are detectable from satellites. Trends of very long-lived gases are usually best measured at the surface. In addition, surface vegetation and soil and oceanic microorganisms are dominant sources of many chemically and radiatively important trace gases, and field measurements are the only way to detect and quantify these emissions.

In the ocean, data from acoustic sounders, *in situ* measurements of the vertical profiles of temperature and salinity, and velocity and other data from buoys are critical complements to satellite data on sea surface topography and temperatures for the study of oceanic heat and momentum fluxes. In addition, *in situ* measurements of primary production, nutrient concentrations, and downward fluxes of organic material are essential for quantitative interpretation of satellite measurements of ocean color and provide the necessary data to examine the very significant role of the oceans in the global carbon cycle.

For the land surface, intensive field studies of major vegetative groupings (biomes) will be necessary to define the water, carbon, nitrogen, phosphorus, and sulfur cycles within these biomes. Such information is essential for interpreting remotely sensed data on areal extent and vegetative density in terms of the global hydrologic and biogeochemical cycles.

Finally, the satellite data on geopotential fields will prove most useful when they are combined with other, land-based geophysical data to decipher deep earth structure and global tectonic processes.

In situ observations for calibration and environmental characterization are equally important in the polar regions. However, because of the hostile nature of the environment as well as the isolation, continued manned measurements are both difficult to arrange and expensive. One answer to this problem is the deployment of extensive data buoy arrays on the drifting ice. In addition to improved measurements of standard oceanographic and atmospheric parameters, observations should also be possible of internal ice stress, ambient underwater sound, water mass movements, and water column properties. Such an expansion in data collection abilities would clearly have to be accompanied by an associated expansion in satellite data transmission capabilities.

The panel emphasizes the importance of a complementary program of earth-based calibration and validation studies and the critical need for sustaining an overall and coordinated program of space- and earth-based studies together with computer-based modeling. If we are to achieve our goal of understanding the earth system, neither space- nor earth-based measurements can stand alone; the two must be developed simultaneously and tied together with models.

AGENCY ROLES AND THE BROADER COMMUNITY

It can be said that remote sensing is no more than a tool; yet it is a revolutionary tool for the earth sciences, offering global, continuous, and long-term data gathered with the same instruments over all the oceans and all the continents. The panel emphasizes that this tool has yet to be fully exploited and notes that this will take decades of future effort by leading researchers. The now-traditional *in situ* techniques required a similar sustained effort from previous generations.

Ease of transition from technological innovation to use, both operationally and in support of fundamental research, requires that the process be treated as a whole, a requirement not well suited to the division of responsibility existing among the federal agencies. Sensor development is usually NASA's responsibility, long-term data collection that of NOAA and DOD, while fundamental research into global systems lies closest to the mission of NSF.

Agency Roles

The lack of a unified approach to remote sensing among the federal agencies is a longstanding issue. To resolve this problem the panel proposes that review and oversight at the interagency level be provided for the programs and budgets for space-based earth measurements of all the agencies involved. This review and oversight activity, following precedents such as the National Climate Program and the National Acid Precipitation Assessment Program, should continue for a decade, have a wide scope, and involve a broader scientific community than that which exists within the government.

There are many concerns about the proper use of remote sensing that could profit from interagency oversight: coordination of agency activities and budgets; use of operational satellites for research; impacts of classification; coordination with surface and *in situ* measurements; data distribution and use; involvement of the nongovernmental science community; and international collaboration.

Our preference is to allow the characteristics of the oversight to evolve from continuous discussions between agencies and the science community rather than to attempt rigid definitions of agency roles. But all agencies, in particular, NSF, must be involved in the oversight process, in recognition of the growing importance of remote sensing to many earth science activities.

The research community also must be strongly represented, even perhaps in a majority; the National Research Council could be an effective channel. We recommend that such a channel be developed.

The Broader Scientific Community

Remote sensing cannot achieve its potential as a research tool until it has been thoroughly used to explore at the frontiers of earth science. Remotely sensed parameters do not arrive at the desk of a research scientist as if they were tables of lengths, masses, or temperatures. Instead, an established scientist is likely to be faced with a long period of self-education during which he or she may achieve few tangible results. Many have faced this choice and decided that it involves too great a risk for their career prospects, even though they are convinced that the use of space data is of great benefit to the subject. Federal agencies can ease this situation.

Specific examples of such support include faculty development awards to allow senior faculty and staff to carry out special study and to go on leave, and to provide promising young faculty with the opportunity to develop facility in the use of the new techniques and data sets. Institutional grants for training of graduate students would also enable the United States to provide the personnel needed to capitalize on this new technology.

In terms of basic research support, we propose that NASA, NOAA, DOD, and NSF work together to set aside funds to support research in areas with a large remote-sensing component and that they signal their intention by periodic announcements of opportunity. Two matters are of primary concern: that funded proposals be of high quality measured against those of the parent community, not merely a subcommunity addressing remote sensing; and that the continuity of funding be ensured.

Restrictions on Data Use

From time to time, especially as data about the earth become more and more accurate and detailed, there are pressures from both national security interests and from the industrial community to restrict various classes of information. The panel notes the need for the United States to map and make available data on the new resource of the Exclusive Economic Zone, and to provide satellite techniques like the Global Positioning System, which has enormous promise for allowing measurements of location and rates of crustal movements that we previously have only been able to infer. At the same time, the panel recognizes the need to restrict data for national security reasons. We urge that such restriction be imposed only after careful thought as to the exact mode of restriction, and that scientists who are familiar with the research needs of the earth sciences community be involved in such decisions. It would be a great national loss if the U.S. research community were to surrender its world leadership for arbitrary reasons of classification.

In summary, the panel recommends that:

- *The space-based earth science programs of the involved agencies be jointly reviewed and coordinated at the interagency level,*
- *The National Research Council be considered as a channel to provide advice on such a program through the appropriate board and committees,*
- *Basic research funds and facilities be made available as required on a peer-reviewed basis,*
- *Faculty development awards and institutional graduate student support be made available for the development of an adequately trained community.*

CONCLUSIONS

Knowledge of the earth is fundamental to humanity's survival. Satellite-borne instrumentation has provided scientific data that, when used in conjunction with *in situ* measurements and theoretical models, has given us a glimpse of the knowledge that could be gained with full application of these techniques. Thus we are now ready to measure, on a long-term basis, the essential processes that govern the state of the earth and its changes over periods of years to centuries.

The first step in this process is the flight of research missions with proven technology to measure those processes not now covered by operational missions, such as ocean physical and biological processes and the global gravity and magnetic field. Together with these research missions we need to build a base for an expanded earth observation effort for the 1990s and beyond. In particular, we need to develop an accurate global technique for measuring rainfall, to provide an imaging spectrometer that will reveal new features of geology and biology, and to develop tropospheric chemical sensors. In addition, we need to continue the existing operational weather satellites, to make the existing data available in a useful form to researchers, and to save our treasure of older data that is now being lost. With all this in place, we will have established the beginning of a global earth measuring system.

The space missions and new instruments are not enough in themselves, however. Adequate data systems must be in place that use state-of-the-art technology for archiving, storage, and accessibility. Major programs of *in situ* measurements including both calibration and validation and process studies will be required; these must be funded at the same time. The data systems and the theoretical models that use the data both will call for the most powerful supercomputers; thus the earth science community adds its voice to that need.

Finally, the panel notes the importance of interagency cooperation in earth science. Collection of data, calibration and validation, archiving and dissemination, and support of basic research cut across many agencies in the federal government. Moreover, the industrial and national defense communities have a real and long-term interest in much of this information. As a consequence, if the United States is to keep its lead in this area of major importance close interagency cooperation is required. The panel recommends that the appropriate oversight be provided at the interagency level.

REFERENCES

The panel has based its recommendations and identification of priorities on relevant reports by committees of the NAS/NRC, especially the reports of the Committee on Earth Sciences of the Space Science Board, *A Strategy for Earth Science from Space in the 1980s and 1990s, Part I: Solid Earth and Oceans*, and *Part II: Atmosphere and Interactions with the Solid Earth, Oceans, and Biota*. The panel has also used the reports of the Committee on Data Management and Computation of the Space Science Board, *Data Management and Computation: Volume 1: Issues and Recommendations*, and *Volume 2: Space Science Data Management Units in the 1980s and 1990s*. These strategy reports are part of a long-term effort by the Space Science Board to set scientific goals and priorities intended to maximize the scientific return on the nation's investment in space science.

In addition, the panel has drawn on the ongoing deliberations of the NASA Earth Systems Sciences Committee, the NASA Earth Observing System's Science Working Group, and the Ad Hoc Group on Remote Sensing for Global Change of the Committee on Space Research of the International Council of Scientific Unions for help in identifying the highest-priority recommendations.

Report of the Research Briefing Panel on Pain and Pain Management

Research Briefing Panel on Pain and Pain Management

Edward R. Perl (*Chairman*), University of North Carolina, Chapel Hill
Benjamin G. Covino, Harvard Medical School and Brigham and Women's Hospital, Boston, Mass.
Howard L. Fields, University of California Medical School, San Francisco
Kathleen M. Foley, Memorial Sloan-Kettering Cancer Center, New York, N.Y.
Wilbert E. Fordyce, University of Washington and University Hospital, Seattle
John A. Jane, University of Virginia School of Medicine, Charlottesville
Thomas Jessell, Harvard Medical School, Boston, Mass.
Donlin M. Long, The Johns Hopkins Hospital, Baltimore, Md.
David U'Prichard, Nova Pharmaceutical Corporation, Baltimore, Md.
Charles J. Vierck, Jr., University of Florida College of Medicine, Gainesville

Staff

Elizabeth Bullitt, *Rapporteur,* University of North Carolina, Chapel Hill
Barbara Filner, *Director,* Division of Health Sciences Policy, Institute of Medicine
Naomi Hudson, *Administrative Secretary*
Allan R. Hoffman, *Executive Director,* Committee on Science, Technology, and Public Policy

We thank The Neurosciences Institute, New York, N.Y., for making its facilities available for the panel meeting.

Report of the Research Briefing Panel on Pain and Pain Management

INTRODUCTION

What is pain? This question has troubled philosophers and scientists since before Aristotle's time. Pain has attributes of a sensation, yet the regularity with which it makes us suffer or feel uncomfortable distinguishes it from touch and other sensations of the body. Is it then an emotion? This uncertainty about the exact nature of pain is reflected in our language. We use the term "pain" for the feeling produced when we are physically injured and also for our emotional reaction when we suffer as a consequence of unkind words or loss of a loved one. Confusion about meaning has contributed to difficulties in research and treatment.

Pain after bodily injury is a physiological phenomenon, the perceptual part of the normal response to a variety of stimuli that are physically threatening to the integrity of tissues, i.e. noxious. Because of the aversive feature of discomfort or suffering, pain as a sensory experience has evident protective value.

Pain develops another dimension when it persists or occurs repeatedly. The discomfort and suffering associated with it disturb the quality of life and can produce complex and profound alterations in behavior. As a consequence, pain is a major health problem. It is the single most common complaint causing people to seek health care. People worldwide take enormous quantities of drugs because of pain. Low back pain alone accounts for well over 10 million physician visits annually in the United States. All of this makes pain a formidable social and economic problem.

THE TWO SIDES OF RESEARCH ON PAIN

Research on pain and its mechanisms has two somewhat different but complementary aims. One seeks to understand the basic nervous mechanisms underlying pain, including their link to behavior and other neurally mediated phenomena. The other investigates the control and management of pain in pathological and clinical situations.

Research on the neural mechanisms of pain has contributed new ideas about how the brain works. For example, the traditional concept has been that the operation of complex nervous systems could be understood largely in terms of the connections or "wiring" of the constituent cells activated by the circulation of brief electrochemical events. Through study of pain mechanisms, it is now known that nerve cells employ chemicals to mediate signals other than those traditionally associated with such function. Study of the newly discovered agents has shown that long-term, chemically mediated interactions between nerve cells are part of the process by which neural actions are produced. Complex control mechanisms by which sensory information is modulated and channeled have also been brought to light.

Improvements in the management of pain have also come about as a consequence of better understanding of its neural basis. Other changes in the therapeutic approaches have resulted from enlightened appreciation of the link between pain and complex behavioral factors. New strategies for pain management include, for example, deliberate stimulation of nervous pathways that are part of the control system for pain mechanisms. Means for controlling pain by drugs while reducing risk of addiction and new categories of drugs useful in such control have been proposed based upon the emerging understanding of the chemical messenger systems in the brain. Also, behavioral testing and behavioral management of patients with chronic pain has become a recognized part of clinical strategy.

Nevertheless, these developments in understanding pain mechanisms and pain control represent only a beginning. The work to date is probably most significant in that it has opened the door to future developments rather than provided definitive answers to the basic scientific questions or to clinical problems.

THE SHAPE OF NEW KNOWLEDGE ON THE NEURAL BASIS OF PAIN

Discoveries about the neural substrates for pain have reshaped our thinking about it as a sensation and altered the direction research in the field will take. Special sense organs for pain have been found and they have been shown to feed more or less dedicated neural pathways. These dedicated networks are paralleled by less specific neural systems whose activity combines input from the pain receptors with other messages from the body's tissues. The neural networks activated by pain-producing stimulation have been found to use a diverse set of chemical messengers to transmit activity and modulate the sensitivity of the nerve cells making up the network. These chemical mediators produce their effects by binding to selective receptor molecules on the surface of target neurons. Moreover, researchers have uncovered a modulatory system whereby the brain alters the flow of activity through the neural networks transmitting pain-related signals.

The practical importance of these discoveries is substantial. Knowledge about specific neural networks for pain has provided a focus for work on pain control. As previously stated, alternate clinical strategies based upon deliberate manipulation of the neural control systems for the networks related to pain are being explored, and information on the neural chemical mediators and the receptor molecules on which they act provides new opportunities for the development of drugs to better control pain while minimizing undesirable effects such as habituation and addiction. This report comments on the general nature of pain as a sensation, the peripheral and central nervous system mechanisms for pain, related behavioral factors, and clinical considerations.

THE SENSE ORGANS FOR PAIN AND CHEMICAL INTERMEDIARIES

There are certain common features in neural arrangements for the various sensations in human beings and other mammals. The starting point is a feature or modification of the environment—a stimulus, e.g., light, sound, tissue swelling. The stimulus must be transduced into nervous system messages. The sense organ where this transduction takes place is more or less specialized to respond to a particular subset of physical and chemical changes.

Sense organs usually are located some distance from the central nervous system. Like most other nerve cells, those leading from the sense organs to the central nervous system have an elongated extension, a fiber, over which the nerve cell activity must be transmitted. In human beings the distance the activity must travel along the fiber can be from fractions of a millimeter to more than a meter. The information represented by the activity also must be conveyed from one nerve cell to the next.

Nerve cells transmit activity across themselves to distant parts by a brief electrochemical event, the nerve impulse. The repetition interval between successive nerve impulses represents the code for information about the activity.

Activity is transmitted between cells at special junctions called synapses, where the membranes of two cells come into close apposition. Commonly, a chemical mediator released by the presynaptic cell (sending cell) diffuses across the tiny gap (about 200 angstroms) to affect the postsynaptic cell (receiving cell) by combining with a receptor specialized to bind a particular chemical configuration. The impulse code and transfer at synapses is how activity generated by the sense organ is conveyed across the various neural elements of the network.

For all of the perceptual reactions accepted as sensations, activity initiated by the sense organ(s) passes through a series of more or less dedicated groups of neurons at a succession of stations in the brain to the level of the brain's cerebral cortex. Ultimately, the activity of these specialized neural networks is translated into conscious experience with its concomitants of discrimination, feature detection, and quantitative recognition.

Until relatively recently, direct evidence for this kind of organization did not exist for pain. The nature of the sense organ, i.e., the transducer, was not understood, although it was known that conduction by a subpopulation of the nerve fibers in peripheral nerves was necessary for pain to be provoked by the usual stimuli. It was also understood that a particular bundle of nerve fibers in the spinal cord, a pathway called the spinothalamic tract, was essential for transmitting information to the brain about pain-causing stimuli. However, knowledge about the origin of this pathway, the signaling characteristics of the neurons that formed it, and the nerve net with which it articulated at higher centers in the brain was sketchy and in dispute.

Sense organs for pain Whether or not there were specific sense organs for pain was a major controversy for over a century, a controversy that dominated theories about the basis of pain. During the past 15 years, a set of sense organs whose activity produces pain has been documented. Earlier difficulties in establishing their existence was largely due to technical problems related to the small size of the neurons involved. In general, pain receptors have thin fibers in peripheral nerves. This special set of sense organs in the body has substantially higher thresholds (requires more intense stimulation) for activation than other sense organs in the same tissues responsive to mechanical, temperature, or other nonpainful stimuli.

Because of their high thresholds, the weakest stimuli exciting the sense organs for pain approach or exceed the intensity that can damage tissue. Thus, sense organs for pain can be considered noxious stimulation receptors or "nociceptors." Skin, muscle, and several other tissues have at least two different types of nociceptors; typically, the nerve fiber of one type is ensheathed by a thin fatty layer called myelin, and one or more others have fibers that lack the myelin sheath.

Nociceptors have been linked to human pain by a convincing experiment. Microelectrodes can be inserted through a human subject's skin into a peripheral nerve and manipulated to record the activity and initiate signals from a single sensory nerve fiber. When the nerve fibers of nociceptors are excited, pain is evoked. When the nerve fibers of other types of sense organs are excited, other sensations are produced, but pain never results.

Much human pain associated with disease derives from subcutaneous tissues such as bone or viscera such as the heart. Relatively little has yet been learned about the relationship between pain and activity of the sense organs of the viscera. Although skin and muscle nociceptors have been soundly documented, the existence of visceral nociceptors has not. Existing information on the cutaneous and muscle nociceptors provides the basis for development of pain-controlling drugs; comparable knowledge about the origins of visceral pain is clearly needed for the heart, intestine, blood vessels, etc.

Peripheral chemical intermediaries The events typically causing pain have as a common basis tissue injury or, in pathological circumstances, changes in tissue of the type produced by the inflammation associated with infection. Inflammation is a complex event involving increases in blood flow, increases in the permeability of blood vessels, and the interaction of certain blood cells with invading organisms or injured tissue. Many chemical substances are known or suspected to be intermediaries in the inflammatory process. These chemicals include amines, peptides, and derivatives of lipids.

Partly because pain is such a frequent concomitant of inflammation and partly from independent evidence, chemical substances have long been suspected to be intermediary between tissue events and activation of the sense organs that evoke pain. A connection probably exists between the kind of chemical events associated with the inflammatory process and either activation of the sense organs for pain or modification of their responsiveness.

For example, arachidonic acid, a breakdown product of a lipid cell constituent, and its metabolites, the prostaglandins and the leukotrienes, appear to play parts in inflammation and allergic reactions. Arachidonic acid and some of its derivatives excite the nociceptors. Furthermore, certain anti-inflammatory agents, such as aspirin, act at least in part by inhibiting the metabolism of arachidonic acid to prostaglandins. Aspirin-like drugs are effective against pain only when inflammation or vascular changes (e.g., headache) are present. Thus, it is possible that prostaglandins are mediators for nociceptor activation only in the presence of inflammation and that other, as yet unidentified, chemical intermediaries are involved in noninflammatory pain.

Testing by chemical agents has revealed important functional differences between the different types of pain receptors in a given tissue. For instance, one type is readily excited by chemical substances that evoke pain (e.g., bradykinin and capsaicin), while another type is unresponsive to the same agents. Such differences imply variations in either the transduction process or accessibility of the sensory terminals to various noxious stimuli. Another functional difference appears after certain kinds of tissue injury or in the presence of inflammation; the sensitivity of some nociceptors may be markedly enhanced while others are relatively unaffected. In fact, once excited, some nociceptors tend to produce persistent activity even though no further stimuli are evident; others do not.

Drug development The recent definitive establishment of nociceptors and the proof of their relationship to pain sensation have provided a far better focus for studies of drug action than has been available in the past. The fact that nociceptors of different types exist and that they have differing chemical sensitivities help explain some of the problems that have existed in drug therapy aimed at peripheral causes of pain.

Pain of peripheral origin, such as that associated with rheumatoid arthritis, is a major health problem. Heretofore, major pharmaceutical efforts to control pain in the periphery have been directed at developing nonsteroidal agents that interfere with prostaglandin formation from arachidonic acid. However, neither traditional drugs such as aspirin nor more recently developed nonsteroidal anti-inflammatory drugs that act on arachidonic acid metabolism are universally effective in suppressing pain of peripheral origin.

There is groundwork now for better-focused research on the basic molecular processes involved in activation of the nociceptors and their relationship to inflammatory damage. Investigation in this direction offers exciting new possibilities for developing drugs of either broader or special effectiveness for refractory clinical pain.

THE PERIPHERAL TRANSMISSION OF PAIN ACTIVITY

Peripheral nerves are composed of the many nerve cell fibers that convey impulses to and from the central nervous system. The nerve impulses are the result of a sequential opening and closing of channels spaced along the membrane that delimits each nerve fiber. When open, the channels allow ions to pass rapidly across the otherwise relatively impermeable membrane; this ion passage is part of the electrochemical event that makes up the impulse. The channels are complex entities consisting of protein and other organic components inserted into the lipid structure of the cell membrane. Blocking the channels blocks conduction in the nerve fiber and effectively disrupts communication. In the case of peripheral nerves, blocking the sensory (or afferent) fibers would cause anesthesia of the part of the body served by the nerve.

One proven diagnostic and therapeutic maneuver in treating certain types of peripheral pain consists of blocking particular nerve trunks with a local anesthetic, such as procaine. Local anesthetics generally work by blocking membrane channels. It has become apparent in recent years that the channels for different classes of sensory fibers may not be the same. Specifically, a number of the thin fibers of the type associated with nociceptors appear to have structurally different channels from those associated with the larger diameter sensory fibers. (The latter conduct the impulses producing bodily sensations of touch, pressure, vibration, and position.) Selective long-term blocking of the thin sensory fibers associated with pain would be an important therapeutic tool for the abnormal pain states often seen after peripheral nerve damage. It would enable other sensations from that part of the body while halting pain. The emerging knowledge about membrane channels and their molecular configuration offers promise for development of agents producing selective and persisting block of nerve conduction.

THE CHEMICAL MESSAGES AT NERVE JUNCTIONS

Study of the chemical transmission between nerve cells is one of the most active areas of neurobiological research. New work has shown that chemical interactions are part of the underlying organizational arrangements in the brain. Much research has been directed at determining how the activity in the periphery is transmitted to the cells of the spinal cord and brain and how it is modified by chemical agents impinging upon neurons of the pain pathways. Drugs working on the nervous system commonly act at neural synapses. This action usually consists of either synergy with a natural synaptic mediator or antagonism of its action. Thus the new knowledge of synaptic mediators and the characteristics of postsynaptic receptors has provided major opportunities not only for understanding the neural mechanisms basic to pain but also for targeting the development of specific drugs.

For example, there has been an explosion of information about neurally active peptide mediators (neuropeptides) and the existence of different, specialized receptor molecules on postsynaptic cells as the crucial link in their action. The neuropeptide story began as a general question about chemical mediators between sensory fibers and central neurons. It quickly shifted in emphasis when high concentrations of neuropeptides were found in the small cells associated with fine sensory fibers and in a particular gelatinous part of the spinal cord (*substantia gelatinosa*). The latter had long been suspected to be the termination point for the fine primary afferent fibers and to be related in some way to pain.

Some commentators and current textbooks erroneously imply that when a neuropeptide is present in a neuron it represents the synaptic mediator used by that neuron. An extension of this idea is that a primary afferent neuron that contained a particular neuropeptide employed the peptide for synaptic transfer to neurons of the central nervous system. On this basis and other indirect evidence, the peptide "substance P" was hailed by some a few years ago as *the* synaptic mediator for at least some pain afferent fibers. This interpretation began to change when it was found that other chemical mediators apparently were present in many neurons that contained neuropeptides. Furthermore, the experimental application of substance P and other neuropeptides produced slower-to-occur and longer-lasting effects than expected for the usual synaptic action.

Reconsideration of peptide action at synaptic junctions has produced a new and important concept. It is now proposed that some peptides with excitatory action function as long-term modulators of neuronal excitability and that the short-term signaling is done by a different chemical transmitter (cotransmitter) liberated by the same synaptic terminal. Several of the neuropeptides also act as hormones, that is, serve as messengers delivered by the bloodstream. Hormonal action fits with the idea of neuropeptides causing longer-term changes in neuronal activity.

Still another aspect of this story is the possible peripheral action of peptides. The sensory neurons with peptides in their central terminals also contain them in their peripheral fibers. Some of the neuropeptides discovered at central synaptic terminals have potent effects on peripheral tissue, including changes of the type seen with inflammation. The part that peptides released by sensory nerve cells play in the inflammation seen after tissue damage or infection needs to be established.

What has been learned about synaptic mediators, the peptides, and their possible cotransmitters is only a first step. We have yet to define the specific actions of many of the peptides and to determine how the receptor molecules they bind to then produce alterations in postsynaptic cells. If peptides are a source of the long-lasting effects produced in the central nervous system in association with pain, modulating their actions with drugs that block the peptide receptors is another avenue to new and crucially needed therapies.

Capsaicin, the substance that makes red peppers hot Part of the peptide history is linked to the ingredient of red peppers (capsaicin) that gives them their irritant qualities. Injected capsaicin both causes pain and, when repeated, has a remarkable desensitizing effect upon cutaneous inflammation and painlike reactions in animals. A few years ago, it was discovered that capsaicin acts more or less selectively upon sensory neurons with unmyelinated peripheral fibers, one class of which are nociceptors. When capsaicin is injected into young animals it causes permanent loss of many unmyelinated fibers in peripheral nerves and a reduction in the number of neurons containing neuropeptides.

The mechanism by which capsaicin produces these toxic and desensitizing effects is not known. However, there is reason to believe that it will be possible to dissociate them through alterations in capsaicin's molecular structure. If this is done, the activity of at least one set of sense organs responsible for pain possibly could be eliminated on a more or less permanent basis. This is still another place where targeted drug development with important clinical use would be possible from further knowledge of the chemical interplay between neurons.

OPIATES AND OPIATE RECEPTORS

Opiates have been used by people for millenia. They have profound effects upon physiologic states mediated by the nervous system, including changes in consciousness, perception, the sense of well-being, respiration, and upon various visceral functions. The most widely appreciated medical credentials of opiates are for controlling pain. The potential for abuse and addiction also is common knowledge. Addiction and habituation (loss of effectiveness upon repeated use) are major limitations in the medical use of opiates for alleviating pain. None of the available opiates that are particularly effective for the control of pain is free of the curses of addiction and habituation. However, synthetic manipulation of opiate molecules has shown that some dissociation is possible between the various effects of opiates on different neural mechanisms.

Opiates, their synergists, and antagonists bind selectively and specifically to a special class of receptors on cell membranes. Several types of opiate receptors have been defined in the central nervous system. Differing affinities for structural variants of opiates suggest differences for the different opiate receptors although their definitive architecture must still be determined. Studies using the binding of chemically defined opiate analogues have demonstrated that opiate receptors are differentially distributed in the brain. In some regions, such as the dorsal horn of the spinal cord and certain parts of the brain stem, opiate-related activity has been implicated in the modulation of pain pathways.

Uncovering the class of chemical mediators for which the opiate receptors are specialized was a major discovery. These natural opiate-like compounds, or endorphins, are proteins and peptide fragments of proteins and thus are chemically very different from the opiate alkaloids, a surprising observation when considering the apparent specificity of the receptors. Endorphins have differing distributions in the brain, and the several kinds of opiate receptors have different affinities for the various endorphins.

The means by which activation of opiate receptors produces biological effects has been only partially worked out. Opiates and their kin in many situations have suppressive or inhibitory actions. Certain endorphins appear to be the naturally occurring inhibitory synaptic mediators in certain neural pathways.[1] Opiate receptors are present in the membranes of certain of the primary sensory neurons, and opiates can block the release of excitatory transmitters from these neurons. The process underlying opiate action includes blocking the ion flux that is required for release of transmitters. In addition to membrane ionic mechanisms, opiates act upon cellular energy systems.

It is common dictum that opiates are analgesics, although many clinicians are aware that the usual doses of opiates do not abolish pain. Recent studies linking work on the sensory component of pain to opiate action have provided a partial explanation. Opiates such as morphine increase the pain threshold for certain kinds of stimuli but not for others. An increased threshold appears for stimuli that preferentially activate a subset of nociceptors with unmyelinated fibers but not for stimuli exciting myelinated fiber

[1]The sensory nerve fibers from the periphery only make excitatory synapses with central cells; that is, their chemical messages tend to make the postsynaptic cell more active. In the central nervous system there are inhibitory synapses as well, in which the activity of one cell suppresses that in another. Inhibitory synapses differ structurally from excitatory ones. They too are usually chemical in nature, with the chemical agents and the postsynaptic receptors different from those for the excitatory junctions.

nociceptors. These observations give insight into why opiate or other drug management of pain has been so difficult. Because several types of nociceptors are capable of initiating pain, each possibly using a different chemical synaptic mechanism, a particular agent (e.g., morphine) will act only on a subset of nociceptor-activated neural pathways.

The work on opiate receptors, in addition to providing some outstanding scientific advances, has also led to new therapies. An example is a method of opiate use for pain control in terminal conditions of metastatic cancer. The discovery of opiate receptors in the spinal cord in association with the central terminations of the primary sensory fibers suggested direct application of opiates to the spinal region. In some cases, opiates now are infused through catheters placed into the spinal canal to provide relief from pain produced by cancerous growth below the level of infusion. This is accomplished without the usual deleterious effects upon respiration and consciousness because the drug is not reaching the upper spinal cord and brain.

Molecular chemistry and the research opportunities New knowledge of the molecular chemistry associated with nervous system action and the neural apparatus for pain constitutes great advances. But, once again, what has been learned is only an introduction. Activation of some and possibly all pain receptors by the peripheral events that cause pain involves a chemical interplay. The full scope of the chemical agents and just how each acts upon the transducer terminals needs to be learned. New classes of compounds that mediate nervous system action have been found, and there is evidence that some of them are associated with the transfer of information between the peripheral sense organs and central neurons. Yet to come is the linking of specific transmitter agents, such as a given peptide, to the nociceptor system and then the determination of the nature of the receptor molecules to which they bind. The opiate receptors are a subset of such molecules, and what has been learned about them gives promise that better analogues and antagonists can be developed. Thus, appreciation of the molecular chemistry of the nervous system elements associated with pain has opened the door for a new and more effective pharmacology. Needed at this juncture are additional basic biological study of the chemical compounds and their actions and definition of their receptors. In this work there is hope not only for pain control but also for effective strategies in managing drug addiction.

THE BRAIN PATHWAYS IN PAIN

Although it is has been known for over a century that a particular pathway in the spinal cord was crucial for pain to be felt upon peripheral stimulation, the origin and terminations of the pathway are recent discoveries. It also turns out that this pathway is only one part of the central connections from the peripheral nociceptors. Additional important progress in defining the neural substrate for pain has come from studying the brain regions to which the nociceptors send their signals.

A rough outline now exists of the neural circuitry for the transmission of information from the nociceptors. Neurons making up part of this system include those in the dorsal horn of the spinal cord, a part of the brain called the thalamus (a major subcortical center for sensory messages) and a particular region of the cerebral cortex. Severing the tract leading from the spinal cord to the thalamic region blocks perception of pain (and temperature) and has been one surgical intervention used to treat persisting pain. However, such chordotomies, when effective, often lose efficacy with time. The reason for this change is not known; it may involve previously unsuspected alterations of connections within the injured nervous system. Now that the importance of the cerebral cortex in pain perception has been established, the basis and significance of interactions between such regions and other brain areas are a fruitful field both for further basic investigation and for study of therapeutic possibilities.

The discovery of specific pain receptors, and of specific pathways within the central nervous system to convey the information they provide, represents a crucial development for understanding pain. With this framework, even with limited details, it is now possible to direct further studies on the workings of the neural networks, their chemical features, and the pharmacology of their manipulation. Until this specific neural machinery was established and some of its features were determined, a systematic approach to pain management was not possible.

Pain modulation by internal circuits of the brain Sensory systems in all mammals are organized so that inputs from different sources interact. A given central neuron has hundreds to thousands of synaptic contacts upon it from other neurons. Reinforcement or inhibition of one input by another can occur. These interactions are the basis of feature detection in vision and tonal recognition in audition. They occur for pain at the spinal and higher levels. Much of the underlying neural connectivity for the interactions and the functional significance of the different elements remains to be worked out. Again, these interactions and the integrations that they represent depend upon chemical mediators, membrane receptors for them, and the activation of intracellular messengers.

A series of pathways descending from higher regions in the brain terminate at points where transfer of information takes place in sensory systems. These descending pathways originate at several levels in the brain and can profoundly influence the transmission of activity related to pain and other sensory modalities. In both experimental animals and human beings, localized artificial activation of certain regions of the brain can abolish pain or pain-related behavior. In part these modulatory actions are produced through the release of particular chemical agents (peptides and other compounds). One site of such action is at the first synapse between the primary sensory neurons and neurons of the spinal cord. Some of the inhibitory effects are actually presynaptic, i.e., they occur on the primary fiber termination itself. One amazing feature of this arrangement is that such a descending inhibitory connection to pain pathways can be activated by opiates acting centrally. These descending modulatory connections are part of the organization that permits a definitive reaction in the face of more than one demand upon behavior. They undoubtedly underlie conscious as well as subconscious control of reactions other than perception and would seem to be a factor operating in manipulations such as biofeedback effects upon muscle contraction.

The circumstances under which the descending pathways are activated in normal life and the implications of the various chemical messengers they employ are not clear at this time. Nonetheless, knowledge of their existence offers another access to control of pain signals through the administration of selective pharmacological agents. What is clearly needed to realize these possibilities is a further unraveling of the neural circuitry, the specific roles of various chemical agents and their receptors, and the way these combine to produce suppression of activity.

Interaction between sensory inputs has been used to control pain by deliberately activating sensory sources and pathways. Electrical stimulation of the skin, nerves, and pathways of the spinal cord have been used with varying success in controlling and modulating persisting pain. Acupuncture is one of the more widely employed procedures involving interactions between sensory inputs, but continuing experience with acupuncture for pain has dampened an early enthusiasm for it as a specific treatment method. Whatever its true value may be, acupuncture is frequently misused as a means of therapy. Proper evaluation of acupuncture and other techniques of sensory interaction in the treatment of pain have been hampered by poor protocols in basic research and by inadequate measurements in clinical investigations and trials. This important area needs rigorous research in the future.

PAIN AS A BEHAVIOR

Pain is a form of behavior. Its report will be influenced by all of the factors that modify behavior. We learn that human beings have pain only if they report it. Information from animals is even less direct. Thus, in contrast to studying neural reactions to noxious stimulation, investigating the perception of pain, particularly in a clinical setting, becomes a study of behavior. However simple this concept may seem, recognition of it was slow to come and represents a major source of enlightenment in clinical management.

A patient's expectations—that is, anticipation or bias—greatly influences the report or reaction. In the clinical situation, curing or healing of the event initially leading to pain does not necessarily end the complaint about it. Persistence of the complaint in these situations does not necessarily mean that the patient is malingering.

Some reports of pain after healing may originate from behavior or other central nervous system activities, but this kind of pain may also have a peripheral basis. An injured tissue can heal in terms of gross repair of damage and disappearance of evident inflammatory reactions; however, pain may persist, perhaps in a modified fashion, because of changes in the sense organs and the central systems they engage. For example, the injury and its repair may alter the physical arrangements of the nerve terminals in tissue and make nociceptors easier to activate than in preinjury circumstances. This could cause pain as if noxious events were taking place even though stimuli or circumstances were innocuous. Objective ways of evaluating such situations are not available and need to be devised. Low back injury and longstanding pain thereafter in the face of apparent healing is a troublesome and important case in point.

Recent attention to mental factors that modify patient behavior in instances of persistent or repeated pain has led to the recognition of special behavioral patterns. "Pain behavior" is a common clinical observation and may profoundly alter the lives of individuals. Unfortunately, evaluation of the mental contribution to the clinical complaint of pain is rarely systematic, and even when done it relies upon imperfect tools. It is known that the issue is important, but protocols for proper evaluation and management are not commonly available. For example, preventive measures such as proper counseling at the time of first contact with the health delivery system could decrease or eliminate many adverse behavioral patterns associated with persistent pain. Systematic implementation of such procedures have not been evaluated to a degree appropriate to the great economic and social burden created by the many persons with "chronic" pain.

THE DILEMMA OF CLINICAL KNOWLEDGE

Information from human experience has been a considerable source of increased understanding of pain and its mechanisms. But there are major shortcomings in the availability and use of such information, partly because of inadequacies in the careful evaluation of the human sensory status. The traditional procedures for clinical sensory testing are time-consuming and poorly quantified. Human sensory evaluation in most clinics is still done using intruments of the nineteenth century. Knowledge from the laboratory on effective means for such evaluations is seldom implemented at the clinical level.

Physicians have had greater concern for loss of human motor function and the obvious incapacities it produces than for losses of sensory capacity. However, many of the most distressing syndromes associated with pain involve sensory deficits, and full information is often missing because of the lack of systematic documentation of the nature of the defects. Protocols and instruments for efficient, effective evaluation of the human sensory status are seriously needed. Insularity of the clinicians and investigators has also contributed to this lack of adequate clinical knowledge; those primarily concerned with neurological aspects or with behavioral problems often give short shrift to each others' domains. Collaboration between those analyzing the neural and the behavioral features in future research is of first importance.

Clinical evaluation of pain management demands careful, controlled trials. Clinical studies of pain therapies rarely are undertaken with appropriate experimental designs and methods of evaluation. In the type of cases represented by back pain, prospective studies are needed to determine the circumstances under which an acute problem of pain in certain patients becomes chronic, persistent, and intractable.

SUMMARY

Research over the past decade or so has provided major new insights on the neural basis of pain and, as a consequence, on certain principles of functional organization in the brain. Now there is a solid basis for well-targeted approaches to further questions on the neural organization and how it may be manipulated to provide effective clinical management of a complaint that represents a major economic and social problem. In particular, the groundwork has been laid for the development of new classes of drugs to control or modify pain when it becomes a health problem. Also, it is now understood that particularly intense or persistent or recurring pain can of itself produce important behavioral and psychic changes. These in turn complicate its management. Progress in understanding and evolving therapeutic approaches to such behavioral and psychic associations with pain have lagged far behind advances in the biological sphere.

RECOMMENDATIONS

The major progress in understanding the biology of pain has come from basic investigations in neurobiology and molecular biology. The classical methods of neurobiology, electrophysiology, and neurohistology have provided a framework. The marriage of these techniques and knowledge with the methods of molecular biology to probe the presence and interactions of complex molecules has been responsible for some of the more dramatic new insights. The advances have come not so much from support dedicated to uncovering the neurology of pain as from general support for research on the organization and functioning of nervous systems. The panel believes that the best way to learn more about the neural basis of pain is to continue to support basic neurobiological research as it has been supported for the past decade. Special encouragement should be given to those interdisciplinary approaches that join the potent tools of modern molecular biology to modern neurohistology and neurophysiology.

There is an important caveat in this recommendation. The operation of the nervous system and such reactions as pain are based not only on the activity of the individual elements and their chemistry but also on a circuitry that consists of the interconnections between a multitude of cells. We are far from understanding all essentials of the circuitry or even the nature of all of the elements making up the circuits. Such information cannot be obtained by working *in vitro*, that is, with isolated tissue or cultured cells. Investigating brain mechanisms requires living animals or viable fragments of the brains of animals. Furthermore, there are major differences in the organization of the nervous system related to pain in different species, and some of these are of crucial importance in extrapolating information to human beings from experiments in animals. In other words, certain essential answers can only come from experiments on mammals, and for some matters only from work on carnivores and primates. Certain other types of experiments on general features of the activity of individual nerve cells and the effects of chemicals upon a general class of neurons may be possible *in vitro* on cultured nerve cells. Clearly, both kinds of research must be fostered with the choice of the species or tissue dependent upon the nature of the question.

The situation for the clinical approach to pain management is less satisfactory than that for research on basic mechanisms. Although advances in the management of pain have taken place, and a better understanding of the approaches necessary has been gained, there are major shortcomings in the use of clinical material for gaining new knowledge about disorders of pain and in applying present understanding to pain management. Overall, the quality of clinical investigation directed at problems of pain is not adequate.

We feel clinical data gathering, clinical management, and evaluation of therapeutic procedures could most effectively be improved by providing clear demonstrations of the appropriate approaches. These could be generated in a limited number of places in this country where clinicians and investigators with the necessary skills and insights happen to coexist. One way to encourage such centers, to focus their energy upon the development of better clinical tools for testing and evaluation, and to improve clinical trials of surgical, pharmacological, and behavioral interventions would be to provide funds dedicated to operating "demonstration centers" for pain research. Five such units at appropriate academic centers across the country would provide a reasonable base for developing appropriate approaches and as sites for training. Good models for diagnostic evaluation and clinical testing of therapies and preventive strategies will be important steps in translating what is known into better care at lower cost.

Report of the Research Briefing Panel on Biotechnology in Agriculture

Research Briefing Panel on Biotechnology in Agriculture

C. Eugene Allen (*Co-chairman*), University of Minnesota, St. Paul
Charles J. Arntzen (*Co-chairman*), The Dupont Company, Wilmington, Del.
Donald C. Beitz, Iowa State University, Ames
Neal L. First, University of Wisconsin, Madison
Robert T. Fraley, Monsanto Company, St. Louis, Mo.
Timothy C. Hall, Texas A & M University, College Station
George H. Lacy, Virginia Polytechnic Institute and State University, Blacksburg
Ronald L. Phillips, University of Minnesota, St. Paul
Peter H. Quail, University of Wisconsin, Madison
Clarence A. Ryan, Jr., Washington State University, Pullman
Ronald D. Schultz, University of Wisconsin, Madison
James H. Tumlinson III, Agricultural Research Service, U.S. Department of Agriculture, Gainesville, Fla.
James Truman, University of Washington, Seattle
Olen C. Yoder, Cornell University, Ithaca, N.Y.

Staff

James E. Tavares, Staff Officer, Board on Agriculture
Carla Carlson, Consultant
Phyllis B. Moses, NRC Fellow
Aida Neel, Administrative Secretary
Allan R. Hoffman, *Executive Director,* Committee on Science, Engineering, and Public Policy

Report of the Research Briefing Panel on Biotechnology in Agriculture

SUMMARY

The accumulated knowledge of science and its application to agriculture have resulted in high-quality agricultural products in the United States and the lowest food costs to consumers of any country in the world. With the advent of biotechnology and the ability to study and manipulate the genes of organisms at the molecular level, limitless opportunities now exist to recombine genes directly and to dramatically increase the efficiency of agricultural production.

The biochemistry, physiology, development, and behavior of animals, plants, microbes, and pest organisms and the genes that control these traits are now more easily manipulated for the benefit of agriculture. Biotechnology enables scientists to study the properties of resistance to disease and to move resistance traits across current reproductive barriers into agriculturally important species. New and highly sensitive diagnostics and improved vaccines can be produced.

The rate of advance in accumulating new information about the basic life processes of organisms, however, is not keeping pace with the rate of refinement of the tools of biotechnology. The ability to study and manipulate organisms using the tools of biotechnology requires broad fundamental information about specific organisms. Much of this fundamental information is currently lacking. Acquisition of new knowledge will depend on an integration of technologies from a number of biological disciplines as well as increased interaction between biology and other sciences. The role of U.S. agriculture as a major positive component in the U.S.-world trade balance hinges on the ability to continue to increase the efficiency of agricultural production and remain competitive in international markets.

Advances in biotechnology that will lead to application and commercialization in agriculture can be greatly enhanced by strengthening basic scientific programs for the study of agriculturally relevant organisms and infusing these programs with the powerful molecular biological techniques that are now available. Special attention must be given to (1) funding and structure of programs, (2) attraction of high-quality investigators to agricultural research problems, (3) flexibility in the regulation of science, and (4) technical resources necessary to conduct high-quality research.

INTEGRATION

Biotechnology is science in transition. It is supported by the relatively rapid development of new techniques that in turn are being used to increase the rate of advance in many areas of science, notably in agriculture. However, the gears of progress in biotechnology have not yet meshed; the rate of advance in fundamental scientific understanding and that of technical capability are not yet equal. The transition to application and commercial development of products derived through biotechnology will be complete when scientists are no longer limited by a lack of fundamental knowledge in the basic biological sciences.

Biotechnology makes use of combined knowledge in basic sciences such as biochemistry, cell biology, immunology, and physiology that is enhanced by an improved understanding of genetics at the molecular level. It is the ability to manipulate cells or whole organisms—through techniques based in molecular biology and genetic engineering—toward a product or application. The use of molecular biological techniques not only evolves from basic knowledge but will further contribute to that knowledge.

In the excitement that has enveloped these newer techniques, many decision makers and scientists have focused on the components of biotechnology rather than on the science as a whole. In separating the techniques from the basic science, emphasis has fallen on the techniques, the tools of molecular biology that allow scientists to isolate a single gene from thousands and then transfer it to a new host. But the power of such techniques provides little advantage for application and commercial development if the basic biochemistry, biology, and genetics are not complete. The tools to isolate a gene can be employed only after a particular gene has been identified and characterized biochemically or genetically. Without the fundamental information on the biochemistry, biology, and genetics of an organism, there is little foundation for biotechnology. There is nothing to manipulate.

The constraint to further progress in biotechnology and its resulting economic benefits is lack of integration. The sophistication of the molecular biology component of biotechnology has outstripped the component defined by a knowledge of specific organisms. To effectively integrate biotechnology as a whole, interdisciplinary research that combines organismal and molecular biology must be encouraged.

The techniques of biotechnology should not be isolated in the organization of research laboratories, in the structure or funding of competitive grants or special programs, or in the minds of decision makers or scientists. The encouragement of interdisciplinary groups that combine the expertise of researchers in basic and applied science with that of molecular biologists is an effective mechanism to bring about integration of biotechnology with the rest of biological research.

Integration promises two important benefits:

1. The quality of biological research will be enhanced through contributions made by classical and molecular techniques, and
2. The rate of discovery will increase, resulting in an expansion of knowledge in the biological sciences and the more rapid development of innovative commercial applications to agriculture and to other segments of the U.S. economy.

RESEARCH OPPORTUNITIES

Molecular biology is becoming successfully integrated with biochemical, physiological, and genetic studies in a few notable research areas. Integration and, subsequently, advances toward practical application in agriculture are developing in these few areas only because the basic knowledge is more complete and permits ready application of the tools of biotechnology. In many research areas directly related to the life processes of agriculturally important species, biotechnology is severely limited by a lack of detailed basic information in biochemistry, physiology, and genetics.

One measure of the rate of progress in the science of molecular genetics is the number of genes that have been sequenced in the past two years. An earlier briefing report[1] noted that fewer than 10 plant genes had beer sequenced by 1982. In mid-1984 the DNA sequences for 86 plant genes and over 1,100 mammalian genes were known. Researchers can now test for correlations between biological properties described at the molecular level and traits desired by the plant or animal breeder for enhanced food production. Significant correlations of DNA sequences with selected traits will lead to an increased use of molecular biological methods in breeding.

The following discussion covers two areas: (1) improvement of production efficiency and quality of crops, food animals, and agricultural products; and (2) protection of plants and animals from biological stress caused by pathogens, parasites, and arthropod pests. Examples illustrate some areas of recent progress, and highlight the potential for improvement and application in many areas of research.

Quality and Efficiency of Food Production

In crop production and forestry, plants can be made more efficient primary producers in ways ranging from net improvements in capturing sunlight for use in the synthesis of carbohydrates, lipids, and proteins to improvements in the quality and balance of these food reserves for animal and, ultimately, human nutrition. Increased efficiency in animal agriculture will come from an improved ability to regulate reproduction, growth, and development toward better feed utilization and increased meat, milk, egg, or wool production. The opportunities to manipulate microbes as symbionts in nitrogen and mineral uptake by plants or as fermenters in the animal digestive tract are great. Engineered microbes are a source of enzymes used to make food products, for example, rennin for making cheese. Microbes can also be used to add value and quality to agricultural by-products.

[1] Report of the Research Briefing Panel on Agricultural Research. Washington, D.C.: National Academy Press, 1983.

Agricultural Organisms

Plants Rapid advances have occurred in areas supported by a more complete fundamental knowledge of plant processes such as photosynthesis. For example, the molecular properties of the plant chloroplast are now beginning to be defined. This self-replicating organelle contains many of its own genes; the cell nucleus contains the remaining genes coding for the chloroplast. Thus, photosynthesis and other chloroplastic functions, including carbohydrate storage and mobilization, are under both nuclear and chloroplastic control. Gene transfer techniques are providing new opportunities for gaining an understanding of the organism and the development of key biological processes including photosynthesis and control of plant growth and development.

For example, the DNA sequences that regulate expression of photosynthetic genes are currently being studied. Researchers have successfully transferred a gene encoding the small subunit of the photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase between unrelated plant species. Genes for the enzyme, transferred from peas into tobacco and petunia, are stably inherited and, very importantly, retain their normal patterns of regulation, by light, and expression in green leaves. Success with the ribulose-1,5-bisphosphate carboxylase-oxygenase gene demonstrates that transfer of genes for important traits between plant species is feasible.

A recent, related development is the cloning of the gene that codes for phytochrome, the major photoreceptor controlling plant growth and development. Phytochrome functions as a switch controlling gene expression in response to light, as in the control of the ribulose-1,5-bisphosphate carboxylase-oxygenase gene. Access to the phytochrome gene will provide opportunities to study the range of plant developmental steps from cell enlargement to flowering by manipulation of this central control switch.

Additional opportunities to improve the quality and efficiency of crop plants are evolving from the study of genes coding for seed storage proteins. Goals are to obtain increased levels of essential amino acids such as methionine in soybean and lysine in corn and to eliminate antinutritional peptides such as trypsin inhibitor and lectin in soybean. The practicality of engineering seed storage proteins was recently substantiated with the transfer of a bean seed protein gene into tobacco. Modified tobacco plants contained more than 2 percent bean protein as part of the tobacco seed storage proteins.

Animals Genetic engineering techniques can be applied to enhance many traits in food animals. Daily administration of mammalian growth hormones, called somatotropins, that are produced by genetically engineered bacteria are being used to study growth in sheep, swine, cattle, and poultry. In dairy cattle the administration of bovine somatotropin increases milk production by 15 to 40 percent and increases the efficiency of feed conversion to milk. This product of biotechnology is being developed for commercial use.

Copies of growth hormone genes have been inserted into swine; the effects are currently being evaluated. Fecundity genes have been isolated in sheep and offer a method to improve reproduction and reduce production costs by increasing the number of offspring in domesticated animals. Two genes regulating the synthesis of enzymes for wool production are being cloned and studied in Australia. Sheep carrying these introduced genes are expected to produce about 25 percent more wool, with little additional energy or cost to production.

Microbes Microbes can be genetically engineered to produce animal growth regulators, diagnostics, and therapeutic substances such as interleukins, interferons, and vaccines. Improvements in food production and processing can be made using genetically engineered microbes. For example, the increased and improved digestion of cellulose and lignin by microbes in the rumen and intestines of domestic animals would permit use of fibrous waste materials, including brush or wood pulp, as feed. Thus, meat and milk might be produced more efficiently, and in areas that are limited in availability of grains. More effective production of essential amino acids and vitamins by the intestinal flora in food animals can lead to more nutritionally balanced and economical diets.

Limitations

The examples cited above demonstrate the power and potential of biotechnology. Practical improvements, however, will depend on a more complete understanding of the biology of major crops and food animals. In plants, few of the enzymes important in plant growth and development have been isolated and characterized. Cell culture and regeneration techniques that support gene transfer methods have not been fully developed for many major crops, particularly corn, wheat, soybeans, and rice. The ability to manipulate agriculturally important plant traits is limited by lack of understanding of the molecular mechanisms underlying these traits and by inadequate molecular biological techniques for gene transfer and plant regeneration.

For animals, further research needed to bring potentially productive results from biotechnology into commercial use includes (1) *in vitro* culture and development of embryos from the one-cell stage, appropriate for gene transfer, to the multicellular stage, appropriate for implantation in surrogate mother animals; (2) regulation and enhancement of gene expression; and (3) adaptation of safe viral vector systems to transfer desirable animal genes into multicellular embryos.

Use of genetic engineering techniques in microbes is currently feasible, but requires continued research in gene isolation, transfer, and testing for efficacy. Further basic research is needed on those agriculturally useful microbes that differ significantly from the well-studied bacterium *Escherichia coli*.

Protection Against Pest Organisms

Diseases caused by bacteria, viruses, or fungi, and damage caused by insects and nematodes are major factors limiting agricultural production. Understanding the relationships between the plant or animal host and these pests can lead to control and treatment. Great advances have been made in the last few years by the application of molecular biological techniques to disease diagnosis and control. However, further research is needed on the basic interactions between host and pest before the newer technologies can be practically applied to protect crops and food animals.

Agricultural Organisms

Animals Hybridoma technology is a dramatic example of the application of biotechnology to disease detection and treatment. The technology permits the production of large quantities of pure, specific antibody from B cells in the immune system against virtually any antigenic substance. Such a monoclonal antibody is already being marketed as an oral prophylactic for calf scours, an enteric disease caused by the bacterium *E. coli*. This common and often fatal infection is responsible for livestock losses of $50 million annually. Monoclonal antibody treatments should be developed to prevent enteric diseases of viral, bacterial, and parasitic etiology in other domesticated species.

In addition, hybridoma technology applied to T cells in the immune system is being used to prepare lymphokines such as interleukins and interferons that increase the immune response of an animal against disease. Lymphokines might provide a method to enhance the immune response of the neonate, the stage of highest susceptibility to infectious diseases and, thus, of greatest economic loss. Monoclonal antibodies or genetically engineered interferons and interleukins should also be developed to control respiratory diseases such as shipping fever, which in cattle alone cause losses in excess of $250 million annually.

Diagnostic tests based on monoclonal antibodies are already on the market for infectious diseases of companion and food-producing animals. Similar specific and highly sensitive diagnostic methods based on monoclonal antibodies or RNA-DNA hybridization are being extended to disease detection and will most likely become routine ways of monitoring for infectious agents as well as for contaminants in meat and milk supplies.

Although studies of the mouse and human systems contribute to an understanding of animal immunology, direct application of hybridoma technology-based diagnosis and prevention techniques to food animals requires a detailed understanding of the basic biology, biochemistry, immunology, and genetics of these animals. To exploit the potential for biotechnology in disease prevention, detection, and treatment, basic biological research on agricultural species and their unique pathogens must be increased.

Plants The opportunities to apply genetic engineering techniques in combating diseases in crop plants are enormous. However, few products are on the verge of commercial application. A significant effort in research on specific interactions between host and pathogen is required before detailed disease control schemes can be designed and developed.

Studies on the mechanism of action of herbicides and of genes specifying resistance to herbicides will increase the options for protecting crops. Genes for resistance to commercial herbicides such as glyphosate, sulfonylureas, and atrazine should soon extend the number of crops that can be protected against susceptible, competing weeds.

Likewise, plant genes that confer resistance to pathogens and other pests can be characterized and ultimately transferred between plant species. Some resistance genes are quickly and easily overcome by pathogens; others have remained effective for more than 60 years. With an understanding of the molecular mechanisms of resistance, scientists will be able to select hardier resistance traits.

Pest Organisms

Biotechnology has been successfully applied to the study and control of viruses and pest organisms. For example, a simple, efficient, and extremely sensitive RNA-DNA hybridization test is now available to identify viroid-free plant material for propagation. Viroids, which consist solely of a small RNA molecule, are responsible for major losses in agricultural crops including potatoes. Several other viral detection systems have been developed using this technique.

Pathogenic bacteria, fungi, nematodes, and protozoan and metozoan parasites can be detected by this technique or a related technique using monoclonal antibodies. Furthermore, recombinant DNA procedures promise to go beyond detection to prevention or cure of crop diseases. For example, RNA sequences complementary to vital messenger RNAs inactivate these messenger RNAs through RNA-RNA hybridization. This strategy could be used to disrupt gene expression in pathogens.

The convergence and exploitation of knowledge from plant pathology and molecular biology is illustrated by the study of the bacterial plant pathogen *Agrobacterium tumefaciens*. Pathogenesis is mediated by a tumor-inducing (Ti) plasmid from *A. tumefaciens* that is transferred into the genome of host plant cells to direct the synthesis of special metabolites needed by the bacteria. Researchers have now adapted this Ti plasmid as a vector for transferring other genes into plants.

As additional molecular mechanisms of pathogenesis become known, prediction of measures that would interfere with many disease processes should be possible. Further research, however, at both the molecular and organismal level of host-pathogen interactions is required. At present, only a few laboratories in the United States are committed to research on the molecular biology of microbial plant pathogens. There is a pronounced deficiency of information on fungi, the major cause of disease in plants, and molecular nematology.

The lack of information on pathogenic nematodes illustrates the importance of integrating basic biological research developments with agricultural research applications. The life cycle, genetics, and developmental biology of the nonparasitic nematode *Caenorhabditis elegans* are being described in extensive molecular detail by several groups that use the organism as a model system to study cell differentiation. The biological information and genetic probes developed for *C. elegans* provide a foundation for studying pathogenic nematodes, which are responsible for $10 billion annual damage to all crops in the United States.

Opportunities to apply genetic engineering approaches to develop new, effective and environmentally safe insect control strategies are promising. Potential advances, however, not only await the development of gene transfer techniques for pest and beneficial insects but also depend upon a better understanding of general insect biology, including sterilization, pheromones, neurobiology, behavior, and insecticide resistance.

For example, the problem of insecticide resistance, which has plagued insect control programs, can be turned to an advantage. Single genes known to confer insecticide resistance, if introduced into mass-reared sterile insects or insect predators such as the wasp *Trichograma*, can permit concurrent use of biological control methods with chemical control programs.

A better understanding of insect neurobiology—its chemistry, metabolism, genetics, molecular biology, and impact on insect development and behavioral patterns—is needed. Effective biotechnological strategies based on more complete neurobiological information could then be designed.

For example, evidence indicates that insect selection of a host plant or animal can be altered by a single gene mutation. Further, direct genetic manipulation of insect behavior should alter courtship and mating activities. In the laboratory the female corn earworm can be induced to produce the pheromone, or chemical sex attractant, of a related species, the tobacco budworm. This manipulation is based upon the known biochemistry of each insect's sex attractant and results in fruitless mating attempts. Additionally, a new brain peptide that regulates the production of pheromones in these and other species has been discovered. Further investigations of specific enzymes and neural peptides can lead to strategies for blocking reproduction by disrupting sexual recognition.

The ability to transfer genes into a variety of insect species will be extremely valuable for pursuing pest control strategies. Gene transfer in the laboratory fruit fly *Drosophila melanogaster* via the P-element vector, a DNA segment that can insert itself into the fly's genome, has been extremely successful. However, this technique, developed for the *Drosophila* laboratory model, has not yet been extended to other insect species of agronomic importance. The Caribbean fruit fly, for example, an agricultural pest indigenous to the southeastern United States, can be used as a model to assess gene transfer techniques developed for *Drosophila.* Techniques developed for control of the Caribbean fruit fly would be directly applicable to future outbreaks of the Mediterranean or Oriental fruit fly. Such an approach could be extended to other insect pest species and genera.

Limitations

Two major factors currently limit the application of biotechnology to disease diagnosis and treatment and control of insect pests in agriculture. First, basic scientific knowledge is lacking on many important biological and molecular mechanisms of disease causation in both animals and plants, and on the molecular biology of plant and animal pathogens and insect pests. Second, biotechnological techniques for gene transfer and expression have not been developed for many pathogenic organisms and are incomplete or suboptimal for others.

These limitations can be addressed by an integrated approach to agricultural science that builds upon the fundamental interconnections between biology and biotechnology. New findings about the biochemical, physiological, and molecular genetic processes of organisms will provide fundamental knowledge that can be applied to improve agricultural productivity. This knowledge is needed to improve the molecular biological techniques essential in improving plant and animal systems in agriculture. In turn, these techniques can be used to advance fundamental research.

ENHANCING PROGRESS IN BIOTECHNOLOGY

The scientific opportunities in agricultural biotechnology are great. Research in several areas—particularly disease diagnostics for plants and animals and therapeutics and vaccines for animal health—has carried basic science into commercialization. In most areas, however, biotechnology must first contribute to expanding the basic understanding of agricultural organisms and their life processes. Combined with this knowledge, use of the tools of molecular biology and genetic engineering can lead to more efficient agricultural production through development of more effective methods and improved varieties of plants and animals. Research advances can further result in spin-off technologies with applications in research and medicine.

Special attention to (1) funding and structure of programs, (2) attraction of high-quality investigators to agricultural research problems, (3) flexibility in the regulation of science, and (4) technical resources necessary to conduct high-quality research is needed to overcome many of the barriers that have slowed progress in agricultural biotechnology.

Funding and Structure

Funding for biological research applicable to agriculture comes from programs in a number of federal agencies in addition to the U.S. Department of Agriculture (USDA). Major resources include programs in biological sciences at the National Science Foundation (NSF), general medicine at the National Institutes of Health (NIH), and biological energy at the Department of Energy (DOE). Current funding levels and patterns, however, are only modestly promoting the integration of newer molecular biotechnologies with basic biological research in agriculture.

Research programs in the basic sciences should be structured to encourage diversity in science and to support interdisciplinary work for the integration of organismal biology with molecular biology. The funding of this research by several agencies, each with a different focus, should be encouraged and expanded.

Diversity

Industry has been increasing its support of basic science through special grant and contract relationships with universities. However, it is likely that private support for university research has reached its peak. Most companies are now building their own internal basic research programs rather than diverting further funds outside to support university research.

Efforts are being made in the intramural research programs of the USDA Agricultural Research Service to foster interdisciplinary research that includes use of the molecular tools of biotechology. A recent National Research Council report[2] discusses these research approaches in detail.

The level of support for extramural basic research that applies to the agricultural sciences varies throughout the federal granting agencies. The size of grants awarded also varies. If one assumes an indirect cost rate of about 40 percent, a grant from NIH pays twice as much per year as the average grant from NSF or DOE and four times as much as the average grant awarded by the USDA competitive grants program. In addition, NIH awards grants for up to five years; the other agencies typically award grants for only two or three years.

Agriculturally important research competes successfully for funding in grant programs outside USDA. However, much of the research needed to advance biotechnology in agriculture falls beyond the mandate of these other agencies. The possibility of continued growth of the USDA competitive grants program is encouraging. But, if these increases are accompanied by a shift in other

[2] New Directions for Biosciences Research in Agriculture: High Reward Opportunities. Board on Agriculture. National Research Council. Washington, D.C.: National Academy Press, 1985.

agencies away from funding basic research with agricultural significance, the effect could damage current advances in biotechnology in agriculture.

Support for basic biology should continue throughout all federal agencies. This structure assures that high-quality research will be conducted from a variety of perspectives.

Special Initiative

Competitive grants are currently the most effective method available to assure—through peer review—that high-quality research is being conducted in high-opportunity areas. The current size of grants awarded, however, is not effective for encouraging interdisciplinary work that combines basic and applied agricultural biology with molecular techniques.

A special short-term initiative for 5 to 10 years is needed to provide the initial momentum for integrating molecular techniques into research programs in various biological disciplines. Special program grants should be awarded at an attractive funding level to interdisciplinary groups of three to five principal investigators whose efforts encompass the fundamental biology of agricultural organisms and molecular biology. This special initiative in basic biological and agricultural research could be structured through NSF and USDA.

Such an initiative, in addition to current programs throughout a number of agencies, could increase in a relatively short time the acquisition of fundamental knowledge and accelerate the application of biotechnology to basic agricultural problems.

Setting Priorities

The application of the tools of biotechnology to agriculture is at an early stage. During this period of research and discovery, it is extremely difficult to predict those areas that might readily yield new and valuable information when challenged with new techniques. Because high-return areas for biotechnology in agriculture cannot currently be predicted, efforts must be made to broadly support research of the highest caliber. High-quality research will lead to more rapid advance in the biological sciences.

Peer review is the best mechanism available for selection of high-quality research. Grants should continue to be awarded through this mechanism; thus, priorities are established by quality rather than by predetermined category.

To reinforce the continuity and focus of biological research in agriculture, a full-time position should be established for a science director within the USDA Office of Grants and Program Systems. Dynamic scientific leadership is needed to establish an integrated research focus for this growing grants program and to ensure that emerging technologies will be used in research on agriculturally important organisms.

Human Resources

Advances in agricultural biotechnology depend primarily on the quality and number of trained scientists drawn to a career in this area of science. Progress will depend upon attracting superior researchers to agricultural research problems. Scientific education should be directed toward establishing a solid foundation in the basic sciences related to agriculture, coupled with an understanding of the power of molecular techniques. This foundation will be critical in establishing strong, competitive programs in U.S. agricultural production and support industries that will successfully meet the increasing challenges of world agriculture.

Training Grants

Training grants and fellowships, awarded on a competitive basis by NIH and NSF, have been effective in enhancing and sustaining research excellence in selected areas of science. The USDA, however, has not placed major emphasis on training young scientists. A study[3] of plant biology personnel and training conducted by the American Council on Education reported that land grant institutions train 70 percent of the postdoctorates and 80 percent of the graduate students in the plant sciences. The 210 universities surveyed also reported a major need for new faculty positions in molecular biology.

Predoctoral and postdoctoral programs in agricultural sciences that offer an integrated approach to training in molecular biology and basic sciences including biochemistry, physiology, and genetics should be expanded.

To equip young scientists with the expertise necessary to advance biotechnology in agriculture, five-year training grants that emphasize the integration of basic science and molecular techniques should be administered by USDA or another agency such as NSF.

Career Development

Outstanding young investigators can be attracted to research that bridges organismal and molecular biology through the availability of career development awards. Such awards would support an individual in an academic research and teaching position for five years at an attractive funding level. Career development awards could be fully funded through USDA, or similarly to the Presidential Young Investigator awards, through matching support by industry or private foundations.

Regulation

Along with the excitement of scientific development and application in biotechnology comes the responsibility of assuring safety and efficacy. In regulation—as in all areas of biotechnology discussed in this report—the barrier to progress in research and application is lack of sufficient understanding of the organisms and their interrelationships.

Basic Research

Regulation can be a barrier to the development of industrial application, because the rules and requirements of regulation have not advanced as quickly as the technology. The opportunities created by biotechnology are accompanied by the need for fundamental research to define guidelines and improve risk assessment. Genetically engineered innovations for agriculture must compete in the marketplace against many cost-effective, established technologies. The required submission of an environmental impact statement prior to each field test—at an average cost to the innovator of approximately $300,000—inhibits application and development.

Targeting new funds toward basic research that provides necessary answers to regulatory questions would be cost-effective over time in advancing technology and protecting the environment.

Reassessment

The Cabinet Council Working Group on Biotechnology organized by the Office of Science and Technology Policy is coordinating the policies of the agencies involved in regulating biotechnology. The interagency working group has stated that current laws are adequate to regulate biotechnology and no new legislation is needed at this time. The working group has also recommended that a scientific advisory panel coordinate the evaluation of scientific methods that support regulatory decisions involving biotechnology.

It is important that regulatory agencies be encouraged to review guidelines and reg-

[3] Andersen, C.J. Plant Biology Personnel and Training at Doctorate-Granting Institutions. Higher Education Panel Reports Number 62. Washington, D.C.: American Council on Education, 1984.

ulations frequently—as often as every 12 months—to ensure that they are compatible with the increasingly rapid advance of the sciences. Regulations that are not at a scientific level comparable with current knowledge and industrial development and that delay testing for relatively long periods could slow research progress in industry and academia and raise development costs to prohibitive levels.

RESOURCES

The impressive power and sophistication of molecular genetic techniques are often matched by the cost of equipment necessary to carry out these techniques. Other resources may be abundant but require assessment or management. These include an array of genetic collections and information on the genes they contain.

Equipment and Facilities

The cost of furnishing a laboratory to carry out high-caliber biotechnology research is approximately $300,000 to $500,000. Equipment such as protein sequencers and DNA synthesizers is expensive but essential. It is also essential that research teams throughout the country maintain well-equipped laboratories to conduct integrated research. Consolidation of efforts in a center of excellence is not necessarily a guarantee of superior research. The award of grants for equipment might be the most effective way to provide for quality research from a variety of groups.

Germ Plasm Maintenance

Biotechnological methods combine genetic resources with molecular analysis. There is an increasing need for access to the fundamental and irreplaceable genetic information that exists in collections housed at universities and state, federal, and international institutions.

As new techniques have emerged for mapping genes, including isozyme mapping and restriction fragment length polymorphisms, the problem of interpretation and coordination of data has become more complex. The present ad hoc efforts to map and coordinate the cataloging of genes in agriculturally important species leaves researchers with incomplete data bases and no mechanism for uniting all available genetic and biochemical data.

Existing resources should be evaluated and organized so that the valuable information they contain can be easily accessed by researchers and used to coordinate research on characterizing the genes of an organism. In addition, management of collections should be reviewed, and the quality of preservation and storage methods should be assessed. If the genetic information of certain plants and animals is lost or destroyed, the opportunity to improve a species by genetic manipulation will also be lost.

Report of the Research Briefing Panel on Weather Prediction Technologies

Research Briefing Panel on Weather Prediction Technologies

Richard A. Anthes (*Chairman*), National Center for Atmospheric Research, Boulder, Colo.
Ferdinand Baer, University of Maryland, College Park
Charles L. Hosler, Jr., Pennsylvania State University, University Park
Cecil E. Leith, Jr., Lawrence Livermore National Laboratory, Livermore, Calif.
Robert A. Maddox, National Oceanic and Atmospheric Administration, Boulder, Colo.
Richard J. Reed, University of Washington, Seattle
Robert J. Serafin, National Center for Atmospheric Research, Boulder, Colo.
Verner E. Suomi, University of Wisconsin, Madison

Staff

John S. Perry, *Staff Director*, Board on Atmospheric Sciences and Climate
Peter Abel, *Staff Officer*, Board on Atmospheric Sciences and Climate
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

The panel acknowledges significant contributions to this report by Drs. Rit Carbone and Joseph Smagorinsky.

Report of the Research Briefing Panel on Weather Prediction Technologies

SUMMARY AND RECOMMENDATIONS

The time is right for a coordinated national effort to improve our scientific understanding and prediction of weather on time scales ranging from less than an hour to more than a month. There is a high level of interest in the United States and other countries in the many exciting scientific questions, such as the generation and decay of convective phenomena including thunderstorms, squall lines and tornadoes, and the interactions of these *mesoscale* phenomena with both larger and smaller scales. The tools—new instruments for remote sensing from the ground and space, powerful computers, and sophisticated theoretical models—exist now and will continue to advance. In addition to answering some of the most challenging questions facing science, this effort will produce large and immediate benefits to society at a low incremental cost, creating especially great potential for high-leverage investments.

To take advantage of the scientific opportunities afforded by the advances in technology, we recommend:

• *High-resolution data sets and the accompanying research to use these data effectively in sophisticated numerical models of the atmosphere* for advances in the understanding and prediction of mesoscale atmospheric systems. Three types of data sets are required:

Continuous Doppler radar data from a network of radars, eventually covering the United States to provide quantitative data of severe convective storms and wind measurements in most precipitation-free planetary boundary layers. These measurements will improve our understanding of energy transports in the atmosphere prior to storm developments and the interactions among various scales of atmospheric motion, information essential to improvement of models and forecasts.

Continuous upper-air wind data from an array of wind profilers (Doppler radar wind measurement systems), also eventually covering the United States. This system will produce observations, nearly continuous in time and up to a height of about 15 km, of the wind, the most important single parameter for numerical weather prediction, and enable major advances in the understanding of atmospheric dynamics and thermodynamics.

Special high-resolution data sets from field programs, such as those to be acquired in the National STORM Program.[1] These research data

[1] The National STORM (Stormscale Operational and Research Meteorology) Program, a proposed major program to improve our ability both to understand and to predict mesoscale weather events, is outlined in *The National STORM Program—A Call to Action*, University Corporation for Atmospheric Research, Boulder, Colorado, 34 pp.

sets are necessary for advancing scientific understanding of mesoscale atmospheric phenomena. This better understanding will lead to improved treatment of physical processes (such as surface fluxes of heat and water vapor, and interaction between clouds and radiation) in both mesoscale models used for short-range weather prediction *and* global-scale numerical models used for medium-range weather prediction.

- *A focused research effort toward studies of several known sources of errors in global forecast models.* These include the tendency for models to develop systematic errors that lead to a model climate different from the real climate, inaccurate forecasting of slowly varying long waves, and inadequate mathematical approximations of such physical processes as radiation and condensation. The necessary research should consist of an integrated effort involving experts in radiation theory, cloud physics and dynamics, boundary layer theory, and applied mathematics.
- *Continued access to advanced supercomputers for both operational prediction and research.* Such computers are essential in extending our theoretical understanding of complex phenomena such as tornadic thunderstorms, to process the enormous volume of meteorological data obtained by remote sensing systems and to run global and mesoscale models of increasing resolution and complexity.

MODERN WEATHER PREDICTION

The atmosphere is a turbulent fluid whose changes encompass an infinite spectrum of temporal and spatial scales. These changes are governed by physical laws, a fact recognized more than 80 years ago when the prediction of atmospheric motion was first formulated as an initial value problem. Since then, weather prediction has evolved from an art into a science, with the quest for improved prediction on all scales of time and space centering on the development of improved numerical models based on an increasing understanding of the physical and dynamical processes governing atmospheric behavior. Continuation of this evolutionary process depends crucially on the availability of two types of data: global observations that are the basis for operational weather forecasting and observations obtained from special field efforts for research purposes. Additional advances in weather prediction require a balanced program of fundamental research that encompasses theory, development, and testing of numerical models, specially designed field programs, and improvement in operational observing, data processing, and communication systems.

Because of the nature of the prediction problem, it is customary to divide forecasts into different types according to time range. We identify four types: nowcasting (zero to six hours), short-range (six hours to two days), medium-range (two days to two weeks), and extended-range and climate (beyond two weeks). Each type has somewhat different requirements; for example, nowcasting (which includes warnings of severe weather) relies primarily on regional observations and extrapolation, while numerical models and global observations of the atmosphere, oceans, and land surfaces play a dominant role in predictions over the longer time ranges. Advances in satellite and computer technology and better understanding of atmospheric dynamics have made it possible to explore the feasibility of extended-range prediction using dynamical models, and some efforts are under way in this direction. This briefing emphasizes opportunities in nowcasting and short- and medium-range predictions, recognizing that progress on these time scales will also contribute to the scientific basis for extended-range and climate prediction.

IMPACT OF IMPROVED WEATHER PREDICTION

Advances in prediction of weather on time scales ranging from less than an hour to more than two weeks are possible, and would provide immediate benefits to the nation. On the shortest time scales, for example, use of a Doppler radar at Denver's Stapleton International Airport during an eight-week experiment in the summer of 1984 conclusively demonstrated the capability to save fuel, property, and life. Aviation authorities determined that the fuel saved due to timely short-range forecasts of low-level winds and attendant adjustments in runway assignments amortized the total cost of the experiment in only two weeks. Even more significant, at least one potentially fatal air crash was avoided as a direct consequence of a microburst wind-shear warning.

The economic and societal impact of severe thunderstorms that carry destructive winds, hail, lightning, tornadoes, and flash floods to all parts of the United States is well known. Some 908 tornadoes occurred during 1984 in 44 of the contiguous United States, causing property damage of at least several billion dollars. In one day, March 28, 1984, an outbreak of more than 30 devastating tornadoes struck North and South Carolina, killing 57 and injuring more than 1,200 people, and causing hundreds of millions of dollars in property damage.

Convective storms are but one manifestation of nature's extremes over the United States. On November 10, 1975, the Great Lakes freighter, the *Edmund Fitzgerald*, fell victim to hurricane-force winds spinning into a rapidly deepening cyclone. This low-pressure area had moved away from the mountains as a relatively benign storm, but then, unexpectedly and spectacularly, intensified as it approached the Great Lakes. Similar cyclonic storms, spawned over the Pacific, in the lee of the Rockies, or off the Atlantic coast, often grow to hurricane-like intensities. Virtually all economic activities, including *agriculture, power, transportation, shipping, state and local government functions, and general commerce*, are adversely affected by these events, which bring heavy rains, flooding, drifting snows, ice storms, high winds, and intense thunderstorms nearly simultaneously to different parts of the United States.

TABLE 1 Scientific Issues in Mesoscale Meteorology

Predictability
Scale Interactions
Forced Circulations
Instabilities
Waves
Boundary Layer and Surface Processes
Air-Sea Interactions
Regional Climate and Climate Change
Severe Weather
Cloud Dynamics and Microphysics
Cloud-Radiation Interactions
Fronts
Atmospheric Chemistry and Air Quality

On even larger scales, seasonal droughts, heat waves, unusually wet seasons, general basin floods, and early and late freezes produce widespread adverse effects almost every year through portions of the United States. Operational temperature guidance forecasts during the extremely cold period of the winter of 1983–1984 were frequently 20 to 30 degrees too warm over large portions of the southern Plains. These poor forecasts had enormously adverse effects on the strategies for operating natural gas pipelines.

The atmospheric science community is ready to undertake intensive studies of the physics of mesoscale phenomena—which are important to all scales of atmospheric motion—focused initially on the central United States, as part of the National STORM Program. The scientific issues to be investigated in the STORM Program are extremely broad, as indicated in Table 1. Research based on new technology will undoubtedly lead to improved understanding of the relationships between mesoscale atmospheric systems and their *larger-* and *smaller-*scale environments. Better understanding of the physics of mesoscale phenomena and the "scale interactions" is required if we are to improve weather forecasting on all four time ranges.

Scientific Basis for Modern Weather Prediction

Development of mathematical models of the atmosphere began in the late 1940s and has advanced in parallel with the development of digital computers. Because of limitations in computer capacity and in the ability to model complex physical processes, the first primitive models dealt with only the larger space scales and neglected the effects of energy sources and sinks. Even so, these early models showed skill comparable to or slightly better than subjective forecasting methods for periods of one to two days. Since then, improved observations, coupled with enormous advances in computer capability, have allowed a variety of physical processes to be incorporated in modern models. As a result, the range of useful global forecasts has improved to at least six days.

Scientists are now developing sophisticated models of much smaller scale atmospheric phenomena. These models are providing new insights into the physical processes that govern such phenomena, laying the foundation for operational weather prediction on these smaller scales.

Medium-Range Prediction

The atmosphere is the prototypical chaotic nonlinear system. This was shown by the simplest atmospheric model, devised by Lorenz over 20 years ago, the starting point for modern mathematical studies of such systems. Because the atmosphere is chaotic, atmospheric models are sensitive to small variations in initial conditions and possess an inherent growth of error. These properties impose a theoretical limit on the range of deterministic predictions of large-scale flow patterns of about two weeks.

Prediction of the global atmosphere by the best computer models is currently accurate enough to be useful for a period of about a week, half the theoretical limit. A decade ago, the useful period was about a third of the theoretical limit. The improvement has come from a better understanding of atmospheric dynamics, models of higher resolution and accuracy, and better techniques for using meteorological data, especially satellite data. These advances, in turn, were made possible by faster computers with larger memories and by the accomplishments of the Global Atmospheric Research Program.

The improvement in medium-range weather predictions over the past decades is best exemplified by the experience at the European Centre for Medium Range Weather Forecasts (ECMWF), currently regarded as providing the most accurate medium-range weather forecasts. ECMWF's success is based on (1) having a very well-focused goal of improving medium-range forecasts, (2) access to the most advanced scientific computers, (3) emphasis on the scientific basis for numerical weather prediction, and (4) an effort to incorporate the best ideas and directly involve leading scientists from many countries, *including the United States.*

Short-Range Prediction

The interaction of all scales in the atmosphere is uncontested. While it is impossible, even with today's supercomputers, for a single model to resolve energetically active mesoscale circulations while simultaneously covering the global atmosphere, there is an increasingly important synergistic relationship between high-resolution mesoscale models that cover limited regions and lower-resolution models that cover the entire earth. The mesoscale models provide important guidance for time periods of about two days into the representation of small-scale physical effects (such as convective clouds) in the global models. Conversely, the global models provide essential meteorological information for the boundaries of the mesoscale, limited-area models.

Further advances in both global and mesoscale models have the potential to improve predictive skill significantly on all scales over the next decade. These advances require *enhancements in quality and resolution of data* together with *more accurate representation of physical processes* such as clouds, coupled with radiation, surface friction, and orography *in numerical models*.

Nowcasting

Predictions over the shortest time scale of zero to six hours include warnings of imminent severe weather and hence have great value in saving lives and averting property loss. Forecasts on this time scale are based on detailed observations and extrapolation of existing weather phenomena, rather than numerical models. Crucial to improved predictions and warnings are better observations from advanced radars and geostationary satellites. These systems provide temporally and spatially continuous observations of severe weather phenomena that almost always are missed by the conventional surface and upper-air observational network.

Although detailed observations from radars and satellites are necessary for significant advances in the prediction of atmospheric systems on short time scales, they are not sufficient. Advanced computers and broad-band communication systems are required to *collect, process,* and *assimilate* the data. In addition, further basic understanding of the structure and life cycle of mesoscale systems is required to enable scientists to *interpret* the wealth of complex information provided by the remote-sensing systems.

CURRENT LIMITATIONS TO SKILL IN WEATHER PREDICITON

Progress in medium-range prediction is limited by major gaps in the global data base, particularly over the oceans; short-range prediction of mesoscale weather over the United States is limited by the inadequate spatial and temporal resolution of the operational observing system. Prediction over all time ranges is limited by inaccurate and incomplete treatment of physical processes in numerical models.

Deficiencies in Data

In spite of modern meteorological satellites, the atmosphere over the world's oceans is still inadequately observed. Forecasts over the United States beyond about a day are adversely affected by errors in the observations made over the oceans. Moreover, data deficiencies are an important cause of the repeated failure of numerical models to predict rapidly developing intense ocean storms that pose a hazard to marine transportation and to coastal installations and activities.

Present oceanic data consist mainly of surface observations from ships and buoys, high-level observations from commercial aircraft, low- and high-level winds estimated from satellite tracking of cloud motions, and low-resolution vertical soundings derived from satellite measurements of outgoing radiation. These components of the ocean observing system fail to provide adequate horizontal coverage and provide almost no data at middle levels in the atmosphere.

Substantial improvement in ocean observations can be made both by expanding the capabilities of current systems and by further development of satellite technologies that have shown promise in preliminary testing. Routine ship and aircraft observations can be increased. Automated balloon-borne instruments, launched from commercial ships and dropsondes from aircraft, can provide needed thermodynamic and wind data. This low-cost concept has already been demonstrated in the North Pacific on Japanese car-carrying ships in cooperation with the Canadian Meteorological Service. Buoys that sense temperature, pressure, and winds can be deployed in data-sparse areas.

The SEASAT satellite, which operated briefly in late 1979, demonstrated the possibility for important advances in observation and analysis over oceanic regions. Particularly impressive were surface wind analyses based on scatterometer data. In one instance, the scatterometer winds (unfortunately not obtained in real time) revealed the existence of an unpredicted, intense storm that battered the liner *Queen Elizabeth II* and sank the dragger *Captain Cosmos*. In addition to wind observations, the SEASAT yielded passive microwave measurements from which water vapor, cloud water, and precipitation in oceanic storms were determined. These variables add a new and promising dimension to ocean analysis.

The required expansion of ship, buoy, aircraft, and satellite observations over the oceans can be achieved only through the cooperative efforts and mutual contributions of the meteorologically advanced nations. By pushing ahead vigorously with its own efforts, the United States can provide important international leadership in developing a substantially improved observational system.

Although the observational network over the continental United States is sufficient for medium-range global weather prediction, it is inadequate in both time and space for detection and prediction of mesoscale weather events. Upper-air data (pressure, temperature, humidity, and winds) are gathered only every 12 hours at spacings of 400 to 500 km (Figure 1). A staggering range of important weather phenomena, from individual tornadic thunderstorms to narrow bands of freezing rain, cannot be detected, much less resolved, by such widely spaced observations.

Until recently, it has been impossible to obtain observations, even over densely populated land areas, with the horizontal and temporal resolution necessary to detect and resolve mesoscale weather phenomena. And even if sufficient observations were available, limitations in communications and in computer power have prevented effective interpretation of the data and their use in mesoscale numerical models. Now, however, with the availability of remote sensing instrumentation, especially satellites and radars, together with the emergence of satellite communications systems and supercomputers, these technical barriers no longer exist, and the science of mesoscale weather prediction is poised for major advances.

Deficiencies in Model Physics

Inadequate treatment of physical processes (such as clouds, radiation, and energy transformations at the earth's surface) in numerical models is a major source of forecast error. Several aspects of these model physics appear particularly important, including *climate drift, energy transfers at the earth's surface, and the dynamics of long waves and blocking patterns*.

Since global models start from observations of the real atmosphere, averages computed over many initial states reproduce well the real climate. But if averages are computed over many 10-day forecasts, one finds instead a model climate differing significantly from the real one. For example, many models develop average temperatures in high latitudes that are lower than observed, leading to a cold bias which results in jet streams that are too fast. This drift of models from the climate of the atmosphere is a source of error in all present-day model predictions.

Prediction models produce incorrect climates because they are missing or misrepresenting important physical forcing mechanisms. Leading components requiring improvement are the treatment of precipitation and associated latent heat release, the interaction of the flow with mountain ranges and the interactions of clouds and radiation. If the *climate drift* could be removed, then the range of useful large-scale forecasts would

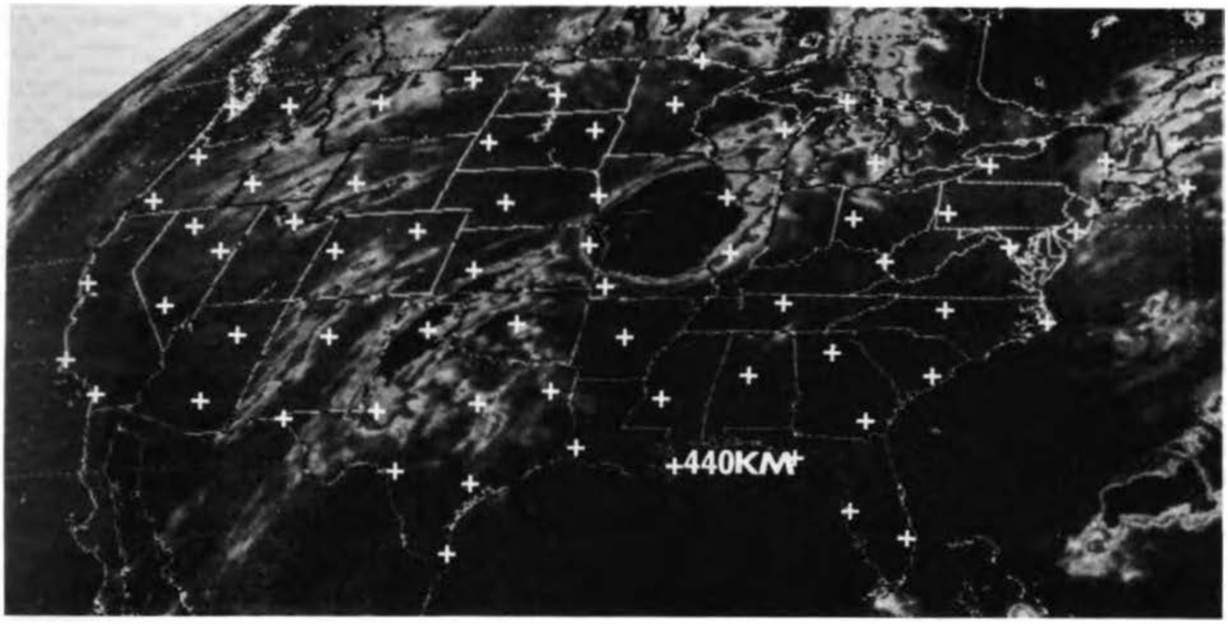


Figure 1 This satellite photograph, taken on 7 June 1982, at 6 a.m. Central Standard Time, depicts a mesoscale complex of thunderstorms covering parts of Missouri, Iowa, and Illinois (the dark, nearly circular cloud pattern over this region). This mesoscale system caused more than 100 million dollars in wind damage. Another mesoscale convective system, a line of thunderstorms, is visible as an elongated dark band over northwest Texas and parts of Oklahoma and Kansas. These common mesoscale systems are inadequately observed by the operational radiosonde network, indicated on the figure by cross marks. The 440 km indicates the distance between the Boothville, Louisiana, and the Apalachicola, Florida, radiosonde sites. SOURCE: National Oceanic and Atmospheric Administration.

be extended by about one day, an improvement of about 20 percent.

There is increasing evidence of the importance of fluxes of heat and water vapor at the earth's surface in modulating atmospheric circulations on all time and space scales, and in particular, in medium-range weather prediction models. These fluxes depend in a complex way on the characteristics of the surface, including the surface reflectivity and roughness, amount and type of vegetation, and most importantly, the availability of surface moisture. Better model treatment of these *surface processes* would undoubtedly lead to better medium-range predictions, and might well be the key to skillful extended-range predictions of temperature and precipitation anomalies.

Another major source of medium-range forecast error is the inaccurate prediction of the slowly changing atmospheric *long waves* and the development, persistence, and breakdown of *blocking patterns*— cyclonic and anticyclonic circulations that persist for a week or more and play major roles in determining temperature and precipitation anomalies. The forecast errors in these large-scale waves are not easily explained, despite the fact that simple theories suggest that the longest waves should have the greatest predictability. Recent theoretical studies suggest that we are close to understanding the mechanisms that determine the evolution of these waves and may be able to improve their treatment in numerical models.

HIGH-LEVERAGE OPPORTUNITIES

Several scientific and technological advances are now converging, stimulating a growing emphasis on mesoscale weather research and forecasting. New remote sensing systems permit observations of the motion of the atmosphere and its heat and water vapor content at much higher spatial and temporal resolutions and with far greater accuracies than previously possible. Advances in computing allow us to consider mathematical modeling of mesoscale weather phenomena for research and operational prediction. Rapidly escalating capabilities in broadband communications technology (including, as a simple example, the transmission of weather forecasts and warnings via cable television), in addition to advances in data processing and display, facilitate the communication of critical data to users.

Implementation of New Technology

Computing has progressed in three important aspects. Large, fast mainframes permit the execution of numerical models that more nearly simulate the physics of the atmosphere. Microprocessor technology permits cost-effective, highly distributed interactive computing essential to the acquisition and real-time processing of complex and diverse data sets. Networking of various computers, through high-speed digital communications, allows distribution of data and analyses for control of large field experiments as well as physical interpretation of model output. Research on both global and mesoscale numerical models requires access to the latest, most powerful computers.

Electromagnetic remote sensing from satellites and from airborne, shipborne, and ground-based platforms permits four-dimensional observations on time and space scales appropriate for both short- and medium-range predictions. These observations span the electromagnetic spectrum from optical to radio wavelengths. The remote sensors most ready for application to prediction are Doppler microwave radars and Doppler UHF VHF radars for the measurement of winds at all levels of the troposphere. As shown in Figure 2, the UHF VHF radar wind profilers provide vertical soundings of the horizontal wind every hour, more than 10 times the frequency of operational radiosonde releases. Passive satellite radiometry at infrared and microwave wavelengths will continue to play a vital role for medium- and long-term predictions particularly over data-sparse oceanic regions, as demonstrated by SEASAT. Combined active and passive microwave sensors aboard satellites will greatly enhance our ability to define patterns of moisture, temperature, and surface winds over the ocean.

Some important observing system requirements cannot yet be met by means of remote sensing. Temperature and humidity soundings with high vertical resolution present particular challenges, but even here solutions are in sight. Highly automated, balloon-borne sensing systems utilizing Omega and Loran-C navigation aids can provide thermodynamic and wind profiles in real time over land and sea. The Global Positioning System, a worldwide navigation system based on satellites, promises even greater capabilities by permitting high-resolution profiles anywhere in the world. A still broader array of ground-based observing systems, complemented by research aircraft equipped with airborne Doppler radar, will be required as part of the next mesoscale field programs.

Field Programs to Provide Special, High-Resolution Data Sets

Special, high-resolution research data sets are urgently needed to document the life cycle of mesoscale weather systems and to initialize and verify new mesoscale research models. While many potential sources of data have been identified as part of the National STORM Program, a few are of particularly high priority because they offer the opportunity to function not only as state-of-the-art research tools but also as permanent observing systems in the operational prediction and warning service.

A *UHF VHF Doppler wind profiler network,*

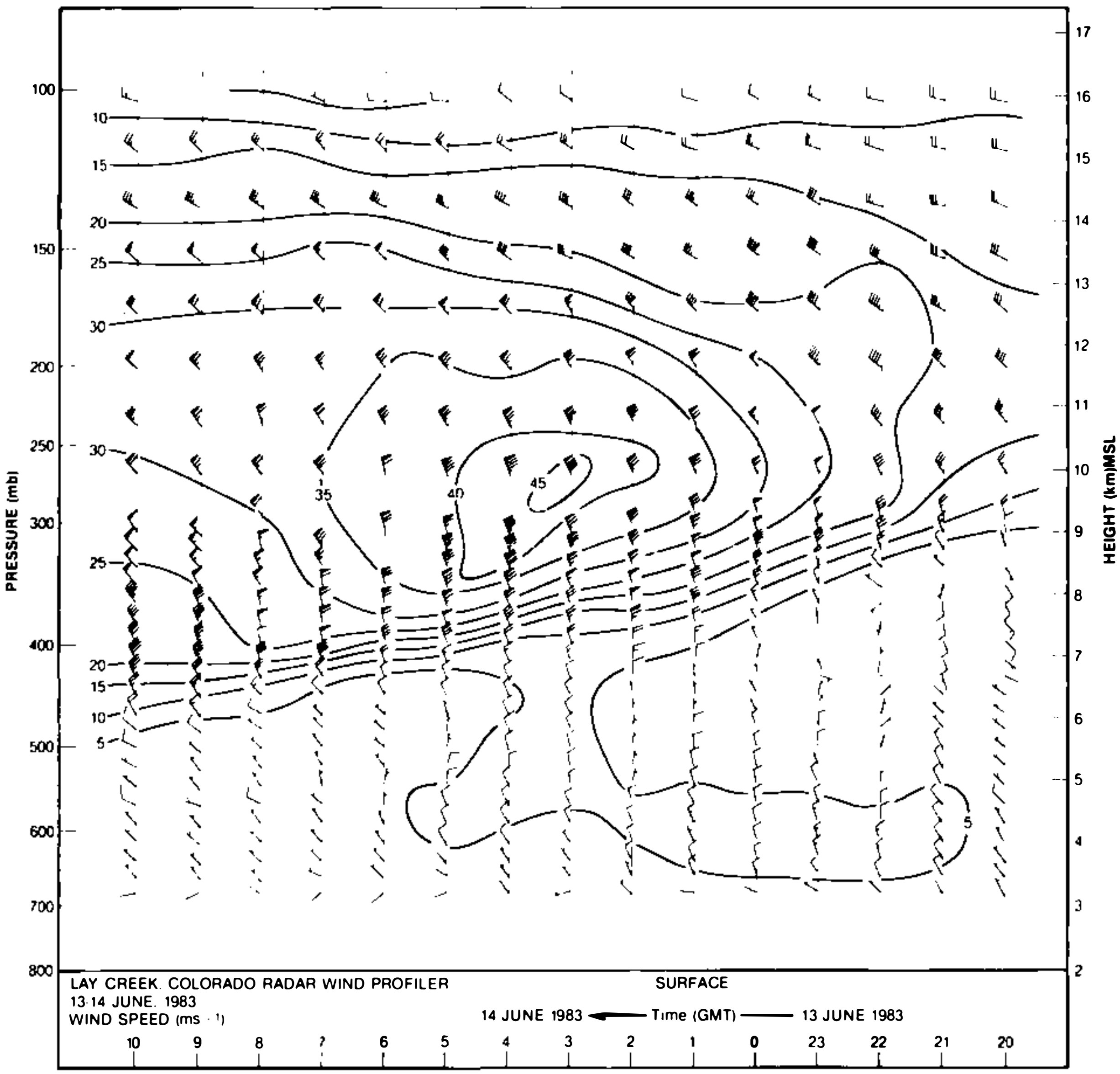


Figure 2 This figure depicts horizontal wind speeds above Lay Creek, Colorado during a 14-hour period in June 1983, as detected by a radar wind profiler. The line at midnight (0 hours) indicates the data provided by the one operational radiosonde launch that would normally occur in this time period. The observational detail provided by the profiler system far exceeds that of the radiosonde. Each symbol, representing one observation, gives wind speed and direction. Note that the jet maximum of $45ms^{-1}$ (~90 mph) would not have been observed at all by the radiosonde. SOURCE: National Atmospheric and Oceanic Administration.

deployed over a large region of the central United States, would provide continuous observations from 1 to 15 km altitude of the wind, the single most important parameter needed for small-scale weather prediction (Figure 2).

A *microwave Doppler radar network*, initially deployed over the central United States, would measure low-level (0 to 1 km) winds in most circumstances, the complete wind field within precipitation regions of storms, and precipitation rates, thereby providing immediate operational benefits in the form of reliable, timely warnings of severe weather including flash floods and tornadoes. In addition, Doppler radars, located at airports, would detect low-level wind-shear hazards faced by military, commercial, and general aviation. These radar data would also contribute to the initialization and verification of mesoscale models for scientific research.

It is essential to build a high level of performance into the U.S. Doppler radar network. Highly sensitive 10-cm Doppler systems with polarization diversity and intelligent signal-processing capabilities would ensure quantitative estimates of precipitation in flash flood situations, hail detection, and the crucial measurement of low-level winds in the clear atmosphere. The NEXRAD design specifications, which have been carefully developed over the past decade, meet all of these requirements, with the exception of polarization diversity, which can be added to the system at a later date at a small incremental cost.

CONCLUDING REMARKS

The time is right for a coordinated national effort to improve weather prediction on all time scales, and particularly on the mesoscale. The interest in the United States and other countries in the exciting scientific questions is high. The tools—new observational technologies, powerful computers, and sophisticated models—exist now and will continue to advance. The potential economic benefit is high, the incremental cost low.

Report of the Research Briefing Panel on Ceramics and Ceramic Composites

Research Briefing Panel on Ceramics and Ceramic Composites

John B. Wachtman, Jr. (*Chairman*), Rutgers University, Piscataway, N.J.
William Rhodes (*Co-Chairman*), GTE Laboratories Inc., Waltham, Mass.
H. Kent Bowen, Massachusetts Institute of Technology, Cambridge
Donald Buckley, National Aeronautics and Space Administration-Lewis Research Center, Cleveland, Ohio
Eric Cross, Pennsylvania State University, University Park
Robert J. Eagan, Sandia National Laboratory, Albuquerque, N. Mex.
Anthony Evans, University of California, Berkeley
Gene Haertling, Motorola Inc., Albuquerque, N. Mex.
Joseph Panzarino, Norton Company, Worcester, Mass.
Dennis Readey, Ohio State University, Columbus
Roy Rice, W.R. Grace & Co., Columbia, Md.
Gerd M. Rosenblatt, Lawrence Berkeley Laboratory, Berkeley, Calif.

Staff

Don Groves, *Staff Scientist*, National Materials Advisory Board
Allan R. Hoffmann, *Executive Director*, Committee on Science, Engineering, and Public Policy

Louis R. McCreight, *Rapporteur*, The Aerospace Corporation, Los Angeles, Calif.
Robert Reynick, *Liaison Representative*, National Science Foundation, Washington, D.C.

Report of the Research Briefing Panel on Ceramics and Ceramic Composites

EXECUTIVE SUMMARY

Advanced or high-performance ceramics, now entering commercial use, are inorganic, nonmetallic materials having combinations of fine-scale microstructures, purity, complex compositions and crystal structures, and accurately controlled additives. Such materials require a level of processing science and engineering far beyond that used in making conventional ceramics. This is true for both areas of primary focus in this report: electronic ceramics and structural ceramics.

Our awareness of the tremendous potential that such materials can play in the U.S. economy and defense has been strongly reinforced by the work that has been carried out by other countries, notably Japan. Successful competition by U.S. industry depends upon economic factors and maintaining leadership in ceramics technology. Promising high-priority research opportunities can be grouped into four broad areas, each of which would support development in both electronic and structural ceramics:

1. New thin films and layer structures with improved properties.
2. Exploration of completely new, multicomponent ceramic crystal structures and composites.
3. The mechanical behavior of ceramics and tough composites.
4. Ceramic processing of large parts and assemblies.

To maintain leadership in ceramics technology the United States needs to strengthen engineering research on ceramics and to increase the number of properly trained scientific and engineering personnel for current and future work. Critical-mass efforts focused on ceramics but drawing upon advances in other disciplines, in advanced instrumentation, and upon computer-assisted modeling are needed.

Such efforts should be centered on ceramics, not dispersed in a variety of scientific and engineering disciplines. Still, other disciplines including chemistry, physics, chemical engineering, and electrical engineering need to be involved. Research should be done in teams of sufficient size and should have a common focus on major themes. This should augment, not displace, existing single-investigator programs in ceramics. The groups should have as a major goal the production of trained personnel. Industrial involvement is essential. However, these group efforts should not be an attempt to perform industrial development or to substitute for necessary, ongoing research and development activity in the industrial sector.

To achieve critical-mass efforts on ceramic research three mechanisms are recommended:

1. A modest number—a minimum of five—ceramic science and engineering programs should be established at universities over the next few years. Each such program should have a focus on aspects of ceramics in keeping with the base of faculty capability on which it is built, but the focus should not be too narrow, and there might be overlap between programs.
2. A summer institute program should be established to bring together leading ceramics researchers.
3. The feasibility of establishing at least one large interdisciplinary ceramics center be studied. Such a center could be equipped to carry out research on the basic science and engineering of ceramic processing and the application of ceramics on a practical scale. It could be viewed as a national facility for interdisciplinary experimental and theoretical work on high-performance ceramics.

INTRODUCTION

Ceramics and ceramic composites are a rapidly evolving category of materials capable of being tailored to have unique combinations of electrical, optical, mechanical, and chemical properties that make them essential and irreplaceable in many engineering applications. Conventional ceramics have both enabled and enhanced many important aspects of our modern technological society; the next generation of ceramics will play a further enabling role for still higher performance devices.

The utility of high-performance ceramics in a variety of electronic, optical, and structural applications is based on their strong chemical bonds and the mixed ionic-covalent character of the bonds. This bond character provides an electrical insulating and semiconducting behavior; optical transparency; ferroelectric, ferromagnetic, piezoelectric and pyroelectric behavior; high hardness and strength; and good corrosion and oxidation resistance.

The utility of advanced ceramics is based also on their ability to play both functional and structural roles, sometimes in combination and often under conditions of high temperatures and severe chemical environments. Functional roles include passive and active components in semiconductor devices, optical devices, motors, transducers, and sensors. Structural roles include applications in which high specific strength, high specific elastic moduli, high wear resistance, and high corrosion resistance or a combination of these properties are needed. Advanced ceramics are also finding increasing use in information processing (electronic and optical), in automation (input-output systems and computer components), in materials processing (including dies and cutting tools), and in engines.

Ceramics are an enabling technology. The competitive performance of many devices and large systems depends on ceramic components that currently make up a small but vital part of the total package. The competitive economic leverage of superior ceramic components is large. These superior components depend, however, on carrying the scientifically demonstrated performance of new ceramics into engineering realization as well as on pursuing the science of still better ceramics. The challenge of realization is a continuum of problems in advancing the science and engineering of ceramic processing to levels of purity, perfection, and scale not previously achieved.

Advanced ceramics is a critical arena of intense international competition. The United States at present is strong in the basic science of ceramics but appears to be increasingly threatened in the extension of this knowledge into the engineering and commercialization of advanced ceramics. Japan, in particular, has a 10-year program in high-performance ceramics with extensive industrial participation and commitment. Japanese firms already control 70 percent of the free-world market for electronic substrates and appear likely to dominate much of the rest of the market for advanced ceramics in the near future unless the United States responds immediately in a more effective manner. Recent events in Europe indicate that similar strong competition founded on excellent technical capabilities is also developing there.

The size of such overseas efforts, their coordinated approach by industry, university, and government, and the close coupling of research and development to applications all underscore the seriousness of the competitive race faced by the United States. Many U.S. firms not previously active in ceramics have identified this as a promising area and have started new programs in ceramics; many firms with existing ceramics programs have strengthened these and expanded their scope. Over 50 firms have joined university-industry programs in ceramics in the last three years. There is a strong commitment to succeed, but also an urgent need to expand the ceramics manpower base if these new initiatives are to prosper and bear fruit. Thus, the time is opportune for government, university, and industrial cooperation in high-technology ceramics.

The key issue discussed in this brief report is the urgent requirement for an increase in U.S. basic science and engineering research for high-performance ceramics. Further technological improvements will result rapidly from the proper application of resources aimed at strengthening and broadening various aspects of basic and applied research in the field. The application of such resources includes not only funding of candidate projects and programs but also directions to ensure a supply of trained professional scientists and engineers necessary to advance the technology.

Designers and manufacturers of advanced engineering systems in the electronics, transportation, computer, manufacturing, and communication fields are becoming increasingly aware that ceramic components offer improved system performance or function. The report focuses on two general categories in which opportunities and growth are expected: *electronic ceramics* and high-performance *structural ceramics.*

ELECTRONIC CERAMICS

At present, electronic applications dominate the world market for high-technology ceramics. Electronic ceramics are essential to the electronics industry, the base of America's current industrial strength. Major applications include low-loss optical fibers, which are causing a revolution in broad-band information transfer; multilayer ceramic-to-metal interconnecting and mounting packages for critical silicon semiconductor integrated circuits; ceramic multilayer chip capacitors (MLCs) of exceptional volumetric efficiency, required to decouple integrated circuits; piezoelectric ceramic transducers for sonar and medical ultrasonic tomography equipment; and chemical, mechanical, and thermal sensors for automobiles and automated manufacture and control.

Ceramics for electronic applications are high-band-gap, largely insulating ceramics and glasses in which a broad range of dielectric, elastic, optical, thermal, electrical conductivity, piezoelectric, and pyroelectric tensor properties can be manipulated and controlled to close tolerances. In most applications the ceramic does not stand alone but is bonded to a metal (for electrodes and conductors) or polymer, which requires precise control and understanding of the interface. Often the metal and ceramic parts are in closely spaced lamellar structures, as for multilayer interconnection packages or multilayer capacitors, and must be coprocessed through the whole ceramic forming and firing process to yield defect-free monolithic hermetic packages.

For integrated circuit (IC) packaging, glass-bonded alumina ceramics are widely used. The green ceramic is usually tape-cast with a suitable organic vehicle, and molybdenum metallization inks are screen-printed for the interconnection wiring. Automation is extensively used in inspection, punching, printing, and lamination. Precise control of shrinkage on firing is essential for the more complex packages. A high degree of perfection has been achieved in this technology. Future-generation IC packages will require ceramics with higher thermal conductivity as the size of the package shrinks to increase computation speed.

Ceramics with lower permittivity are urgently required for advanced electronics. One of the simplest ways of lowering the permittivity in a ceramic or glass is to introduce a void space with a relative permittivity of unity. There is a real need for new processes and design methodologies to introduce such void networks without serious deleterious effects on mechanical strength.

Current technology, with the exception of optical fibers, is generally an extension of standard ceramic technology—i.e., powders pressed or formed with binders and sintered to densify ceramics. Results of recent research on solution processing have improved the purity and size distribution of powders and have refined the microstructural control required to attain desirable electrical and mechanical properties. However, the present technology is crude relative to that used to process semiconductors.

Major processing advances should also center on developing the underlying science and technology to make lower-temperature pinhole-free films in the 1 to 25 micrometers thickness range. These are essential for the evolution of the next generation of multilayer packaging and multilayer capacitors. An evolutionary change to ceramic processing for packages using organically derived ceramic films has started. An intense interdisciplinary effort involving organic chemists, ceramists, metallurgists, and physicists is needed to understand and exploit this emerging technology. At present, existing ceramic systems (e.g., Al_2O_3) are being explored, but to meet requirements for low dielectric constant, high thermal conductivity, high strength, and matched thermal expansion to the semiconductor, new materials must be developed.

A revolutionary change in electronic ceramic processing will occur when semiconductor processing is applied to the ceramic components. Approaches to control the physical properties of ceramics on or near a chip, such as selective area ion implantation and laser-induced recrystallization, will greatly enhance packaging capabilities. Related studies to improve the crystalline perfection of thin-film structures will make possible the fabrication of high-permittivity ferroelectric and pinhole-free thin films for capacitor applications.

Barium titanate ($BaTiO_3$) is the dielectric in most current multilayer ceramic capacitors. Lead and bismuth borate glasses are frequently used to lower firing temperatures and to permit the use of less expensive silver and palladium alloys for internal electrodes. Tape-casting is widely used for green-forming. Precise processing is required because commercial dielectric bodies must often be fired to a specific nonequilibrium phase assemblage to achieve the permittivity and temperature characteristics required for application.

Electrically active ceramics (e.g., ferroelectrics and varistors) require precise control of microstructures and chemistry to attain suitable properties. Although the physics of varistor and ferroelectric behavior is not fully understood, performance has been significantly improved through process control. For example, the switching field for varistors can be increased to 120 kV/cm from the 5 kV/cm found in commercial devices by using ceramic precursors derived from solutions to control grain size and maintain a very fine-grained structure.

In *sensor applications*, electronic ceramics fulfill many different roles that make use of the unusual property combinations achievable. For force, stress, and vibration sensors, the polarized piezoelectric ceramics in the lead zirconate titanate (PZT) and related families and the newer electrostrictive materials in the lead magnesium niobate relaxer compositions are currently of major interest. In a range of chemical sensing applications, the semiconductor ceramics in the zirconia and titania material families and zinc oxide currently are very important. For oxygen-sensing and pressure control, zirconia (ZrO_2) and titania (TiO_2) are useful. Furthermore, in the titania- and zinc oxide-based systems, specificity to a number of hydrocarbon gases can be controlled. New advances and their integration with semiconductor technology ensures a bright, expanding future for these systems.

Sensors based on ceramic systems are in a dynamic research and engineering phase. The balance between sensitivity and selectivity needs to be addressed through fundamental studies in such areas as catalysis, as well as through clever engineering designs to build in film sensors of varying sensitivity or selectivity on one chip. The number of sensors in the automobile, medical, and home market will grow markedly, and many will be ceramic-based because of the inherent stability of this class of materials.

Current *optical fibers* are silicate glasses that contain only parts per billion levels of impurity. Precise control of the refractive index profile is achieved by the addition of germania in the preform. Processing, in this instance, has kept pace with theoretical development. Absorption coefficients are near their theoretical limits for silicate fibers, but research on longer-wavelength fluoride-base glass fiber is in its very early stages. Fluoride glass fibers are predicted to give an order of magnitude improvement in performance, but the knowledge base of fluoride glasses is still limited.

It has been suggested that the new ability to use optical fibers for high-density, high-bandwidth communication must eventually force a change from electronic to photonic signal processing. In this revolution, electronic ceramics may play a vital role in providing the materials base for the simpler hybrid technology that will be required before full integration is accomplished in gallium arsenide and related semiconductors. Research is needed to develop existing families of perovskite, tungsten-bronze, and lithium niobate electrooptic materials to the level where deposited or in diffused optical guiding structures can be reproducibly fabricated.

In a more sophisticated area, holographic signal processing can be achieved, using the photorefractive effect, in electronic ceramics such as $LiNbO_3$, $KNbO_3$, and $BaTiO_3$. This effect has application in interconnection control as well as more unusual possibilities in nonlinear coupling for two- and four-wave mixing and signal up-and-down frequency conversion.

The future promises continued dynamic growth in these areas, together with burgeoning new families of applications in hybrid optoelectronics that use ceramic components in optical waveguides; in electrooptic and acoustooptic modulators and switches; and in sophisticated nonlinear optical configurations for four-wave mixing, phase conjugation, and optical frequency multiplication. However, to realize the potential in these new areas it will be necessary to broaden the base of ferroelectric single crystals that can be made available in highly perfect single domain state. To move to more complex compositions so that properties can be adjusted to desired values one must learn new techniques for the growth of new perfect crystals in new solid solution compositions.

STRUCTURAL CERAMICS

High-performance structural ceramics uniquely combine high strength, strength retention at high temperatures, high hardness, dimensional stability, good corrosion and erosion behavior, low coefficient of friction, high elastic modulus, and low mass density. Structural ceramics are used as coating, monolithic, and composite components. Major applications include tooling for metal working; wear components in a variety of abrasive environments; bioceramics for bone or tooth replacement; and military ceramics for radomes and armor.

Major efforts are under way to apply structural ceramics to automotive reciprocating engines for wear components, turbochargers, and a variety of diesel engine components; gas turbines for regenerators, recuperators, and stationary and rotary components; and ceramic bearings for both rotary and stationary elements. Success in any of these future applications will dynamically affect market growth, improve energy efficiency and the balance of trade, and reduce dependency on critical metals. The combination of cleaner burning by gas turbines and greater fuel efficiency for any heat engine using ceramics would improve environmental air quality.

Monolithic structural ceramics are currently based on silicon nitride, silicon carbide, partially stabilized zirconium oxide, or alumina systems. Processing and microstructural advances have made significant strides in improving mechanical properties and reliability. Considerable research effort is still needed to refine further these parameters before there is wide-scale utilization of such systems in structural ceramics. The ambient temperature brittle fracture behavior is well understood from a fracture mechanics perspective. The strength-limiting flaws have been identified, and work is under way to either shrink their size or eliminate them through improved processing; this work must continue in order to enhance reliability. Improvements are needed in powder synthesis, powder properties, near-net-shape fabrication methods, microstructure control, mechanical properties, and nondestructive testing methods. High-temperature fast and slow failure mechanisms due both to mechanical and to chemical environments must be understood and controlled. It is possible that new alloys of existing matrix materials with improved properties can be developed. Major efforts are required to realize the full potential in long-term high-temperature applications.

The existing structural ceramics for high temperatures are very useful but offer a limited range of choices. At intermediate temperatures the range of choices is significantly broadened by the availability not only of other binary compounds but also of a variety of ternary and higher compounds. These latter are almost exclusively compounds of silica with various lighter metal oxides such as MgO or Al_2O_3. For thermal structural applications, lower stiffness and especially thermal expansion are key parameters to control for wider applicability. The extensive use of these silicates stems from the existence of most of them in nature but is also based on one or more properties or a combination of properties giving superior performance over binary oxides. This superiority stems from the greater range and complexity of crystal structures that commonly accompany a greater number and diversity of atomic constituents. The diversity of structures and resultant properties is not limited to thermal structural applications but may also find applications based on unique properties. New refractory oxynitride and oxycarbide glasses are under study, and these may broaden the range of properties available for monolithic glasses or as additions to existing materials for liquid-phase sintering.

The basic research components for identifying new, more complex ceramics are predictions of phase relations, crystal structures, and key properties of interest. The greatest need for estimation and guidance is in areas such as thermal expansion and high-temperature solid-to-solid phase transitions. Important criteria are identification of materials having (a) low expansion (e.g., carbosilicides and nitroborides) for improved thermal shock-stress resistance; (b) high expansion for coatings more compatible with metal substrate expansion; (c) high thermal expansion anisotropy for use in composites; (d) phase transformations useful for phase transformation toughening; (e) easily sheared structures for lubricity; (f) natural fibrous structures for toughness (e.g., more refractory analogs of jade); and (g) more refractory glasses for sintering aids, bonding, fibers, and bulk glass applications.

Structural ceramics for *wear and bearing applications* depend on an understanding of their fundamental tribological properties. Some measurements of adhesion, friction, and wear behavior primarily of monolithic ceramics have been conducted, and limited studies have been made of the effect of environmental constituents on tribological behavior. Surface studies have examined some interfacial phenomena for ceramics in solid state contact with other ceramics, with metals, and to a limited extent with polymers.

The fundamental tribological properties of ceramics must be related to other physical, mechanical, and chemical properties to provide a basis for material performance. Interface studies of ceramics in contact with themselves and other materials need to be expanded so that theoretical models can be developed to predict interfacial behavior.

There is little understanding of conventional fluids for the lubrication of ceramics. Liquid lubrication technology has been developed primarily and nearly exclusively for metallic systems. The mechanism of lubrication of ceramics must be examined.

Coatings, ion plating, sputter deposition, and chemical vapor deposition should be investigated for improving the tribological performance of ceramics. The interface between coating and substrate is critical because, under the high contact stresses encountered in mechanical components, failure at this site limits coating life. Surface treatment of monolithic parts or coatings by ion implantation, laser glazing, and electron-beam treatment should be investigated to improve the tribological performance of ceramics.

Thermal barrier coatings for applications such as adiabatic diesel and gas turbines are important and warrant further research and engineering to optimize long-term performance. New processes may be required to fabricate these coatings economically with the required microstructural perfection for long life. Little is known, for example, about the effect of microstructure and interfacial properties on high-temperature environmental and mechanical fatigue.

Ceramic matrices combined with particulates, whiskers, or fibers of a different ceramic compound or metal for enhanced performance have yielded *composites* with five times the resistance to fracture (toughness) of monolithic ceramics. While ceramic matrix composite research is embryonic, it has a bright, but uncharted, future. Composites could find application in lightweight, stiff, dimensionally stable space structural members, as key building blocks for new high-temperature rockets, or in many of the applications discussed earlier.

Ceramic matrix composites toughen by increasing the fracture energy of the material system over what can be achieved in monolithic ceramics. Toughening is achieved through crack deflection around the fiber (whisker or particulate), the increased stress necessary to break the fiber, or the energy required to pull the fiber from the surrounding matrix. Typically, a well-designed composite will fail noncatastrophically under the highest loads. The interfacial strength of the matrix-reinforcement phase is a key parameter to adjust for optimal toughening. Significant advancements in recrystallized glass reinforced with silicon carbide or carbon fibers have been made by controlling the interfacial bond strength. A number of other parameters are crucial in this field, such as thermodynamic stability, thermal expansion mismatch, relative elastic moduli, processing protocols, and microstructure.

New research is needed on quantifying the improved mechanical properties for composites, particularly fracture energy or toughness, as the fracture mechanics routines for monolithic ceramics are invalid. Literally thousands of combinations of matrix and reinforcement phase are possible. An empirical approach has produced some promising candidates, but this is not recommended as a sufficient approach for the future. Instead, a judicious selection of systems should be based on sound scientific principles, including thermodynamics, and relevant physical properties. It may be necessary to collect fundamental thermodynamic and phase diagram knowledge before selecting a high-temperature-compatible reinforcement phase. Once chosen, new fibers and/or whiskers may be required. Interfacial properties between phases need to be investigated and perhaps modified through coatings to adjust for the desired level of bonding. Processing composites to form dense, flaw-free, uniaxially aligned and, alternately, orthogonally aligned composite structures requires major science and engineering efforts. Quantitative microstructure characterization and iterative feedback through the synthesis, processing, and properties loop will lead to optimized structures that might revolutionize such systems as the gas turbine, rocket nozzle, or radome technologies.

Designing for Ceramics

Ceramics present special requirements for successful application under high stress. They characteristically undergo brittle failure and may show creep or slow crack propagation at high temperatures or under reactive conditions at stresses somewhat below the fracture stress. Stress concentrations can cause catastrophic failure. A new field of designing from brittle materials, including ceramics, is taking shape. Aspects include computer-assisted analysis of the stress distribution in great detail, probabilistic design based on statistics such as the Weibull statistics of failure, and time-dependence of mechanical properties of materials under load. Design should also take into account the capability of processing large and complex shapes and the impact on properties of scaling up to large sizes. Joining of ceramic parts into assemblies also presents special problems for both design and processing.

Processing of Ceramics

The theme of ceramic processing runs through the discussion of both electronic and structural ceramics. Improvement in performance requires improvement in the science and engineering of ceramic processing. Promising directions include the use of ultrafine powders and the use of chemical routes that supplement or bypass some of the powder processing stages. Special processing requirements include the processing of fine-scale layer structures, processing of ceramic composites, joining of ceramic parts, and processing of large and complex parts to near net shape to reduce final machining requirements. Fundamental processing studies would greatly benefit from a connection with the engineering processing of large parts. Several mechanisms can occur in parallel in ceramic processing and may obey different scaling laws. In addition, large parts frequently have thin sections so that modeling and optimization of processing based on small-part fundamental studies need the test of comparison with experiments on large, complex, and costly parts.

Ceramic processing technology is relatively attractive from the point of view of raw materials supply and hazard control. Most of the required raw materials are domestically abundant. Solid ceramics in finished form are among the least hazardous materials.

Hazards associated with some powder, fibers, and vapors occurring during processing are regularly addressed by industrial safety engineers. Studies show such hazards can be controlled by proper process design.

EDUCATIONAL CONSIDERATIONS

Ceramic science and engineering as academic disciplines are concerned with the understanding of structure, properties, processing, and applications. Thus, ceramics is multidisciplinary learning from the principles provided by physics and chemistry, and the principles of chemical engineering, mechanical engineering, electrical engineering, etc. These principles are invoked through the continuum of ceramics processing, structure, and properties, with the chemical sciences playing a larger role in the processing and structure relationship and with physics and the other disciplines being more important in relating structure and properties.

There are 12 institutions certified by the Accreditation Board for Engineering and Technology that grant B.S. degrees in ceramics or ceramic engineering. There are 15 schools that offer formal advanced degrees in the field; many more schools granting degrees in "materials science" or "materials engineering" have faculty performing research on ceramics. Approximately 300 B.S., 90 M.S., and 35 Ph.D. degrees are awarded annually in ceramics or ceramic engineering.

At present, a large fraction of the B.S. graduates enter the more traditional areas of the ceramics industry such as glass and refractories. An increasingly larger fraction of the B.S. graduates, perhaps 40 percent, currently enter the electronics field. On the other hand, virtually all the M.S. and Ph.D graduates are filling positions related to high-technology electronic and structural ceramics. The demand for advanced-degree graduates, particularly Ph.D.s, is far greater than the supply. As a result, many positions that could and should be filled by ceramics graduates are being filled by graduates from related fields, such as metallurgy, geology, and chemistry, who are being retrained during their first few years of industrial experience.

As high-technology ceramic production expands, the demand for B.S. graduates will increase. By 1990, most B.S. graduates will be entering the high-technology industries. At present, probably more than half of those entering graduate ceramics education have an undergraduate degree in the field. Thus, the number of B.S. graduates must grow to meet the needs of both expanding industry and expanding graduate research.

The livelihood of many of the undergraduate programs depends critically on the coexistence of a strong externally funded graduate program within the same department. A significant fraction of any new external support should be invested in those schools having both strong undergraduate and graduate programs to ensure that a continuous supply of high-quality B.S. graduates is available for both industry and graduate study.

Attracting new faculty and more students of high quality into ceramics programs in universities requires strengthening of these programs. Critical new efforts that permit the student to be a part of a balanced program supported by involvement of other disciplines and the use of modern equipment would be most important.

CONCLUSIONS AND RECOMMENDATIONS

The general conclusions reached in this study are:

- High-performance ceramics represent a class of materials that are essential for many electronic, optical, and structural applications in industry and defense. Such materials, which include monolithic ceramics, ceramic composites, and coatings, will find an expanded future use in such applications.
- The United States has led in basic research and development of products in high-

performance ceramics. However, this position is shifting because of the strong challenge to the United States from Japan and, to a lesser extent, from Europe. The primary aspect of this challenge is their ability to transfer fundamental research in high-performance ceramics to engineering practice and commercialization rather than their superiority over this country in basic research.

- Engineering research as well as basic research on high-performance ceramics must be strengthened and given needed impetus to ensure U.S. scientific, technological, and engineering leadership in the field. This calls for a large and appropriately coordinated effort on the part of universities, industry, and the federal and perhaps state governments.
- Preparation technology for improved thin films and layer structures designed for electronic and optical applications (e.g., substrates or capacitors), mechanical applications (e.g., friction and wear reduction), and chemical and thermal protection could greatly benefit from advances in ceramic processing and property science. The preparation science of layered ceramics and the study of surfaces, interfaces, and joints require interdisciplinary efforts by ceramists with other scientists and engineers.
- Completely new, multicomponent ceramic systems offer the potential of higher levels of performance as electronic or structural ceramics. Advances in chemistry, including chemical modeling and computation, offer improved predictive power as well as the ability to handle complex systems.
- The mechanical behavior of ceramics and tough composites for both electronic and structural use, including the behavior of ceramic fibers and interfaces, is an area of rapid progress. The identification of flaws as well as other high-temperature failure mechanisms needs further study, quantification, and the development of means to eliminate or control them. Toughness characterization, life prediction, testing, reliability assurance, and design with brittle materials must be improved.
- Interdisciplinary research offers major strategic gains on the complex problems of advanced ceramic processing and application, which must be solved as a materials system to supplement the knowledge gained by study of the isolated components. The components range from raw materials preparation, forming, sintering, joining, and fabrication to nondestructive evaluation, and include all aspects of design for current and future high-performance requirements for both electronic and structural ceramics applications. Novel chemical routes that combine or bypass stages of conventional ceramic processing also need in depth study.

No single university, government laboratory, or industrial organization has all the elements of what is foreseen as the necessary critical mass of multidisciplinary personnel to undertake the required programs. There are indeed several establishments that have such a grouping of talented personnel, and a very small number of others are in the early stages of development. However, neither any single U.S. organization nor the total of all pertinent U.S. organizations in these subjects currently has the dedication and focus of the Japanese efforts in these fields.

The following recommendations are offered for consideration:

- The United States should establish an initiative based on various programs designed to expand ceramic science and engineering research at universities. Such programs would be so configured as to ensure a reservoir of more trained graduate students, as well as a greater number of critical-mass size groups involved in ceramics research at universities. The bulk of the support for these programs could be provided by the National Science Foundation (NSF) and other federal agencies with the remainder (say 40 percent) coming from industrial, university, state, and other matching sources. A substantial portion of the funding in the early years should be for major equipment items. Several of these programs should be established to strengthen the ceramic education infrastructure at universities that already have existing strong undergraduate and graduate programs in ceramic science and engineering.

• A "summer institute" program should be established. In this program a number of leading ceramic researchers could gather for about a month and discuss their work, set priorities, make recommendations to NSF and other agencies for future funding, and coordinate efforts. Different groups concerned with different aspects of ceramics could rotate annually, with perhaps a core group of individuals. University, industry, and materials laboratory personnel would be included.

• The establishment of at least one large interdisciplinary ceramic center should also be considered. This center could be equipped to carry out research on the basic science and engineering of ceramic processing and the application of ceramics on a practical scale, including the relation of processing variables to specific model parts. For example, it could include experimental capabilities for detailed in-process studies of flows and reactions occurring during processing and with computer support (perhaps including access to a supercomputer) to permit detailed modeling of processing. Visits from scientists and engineers could be encouraged and provided for in the budget. The center could be viewed as a national facility for interdisciplinary experimental and theoretical work in electronic and structural advanced ceramics. It would contain an internationally recognized strong nucleus of permanent staff who would attract on a temporary basis individuals from industry, universities, and the national laboratories. Such a center would require funding, which could be jointly provided by federal agencies, industry, and perhaps state governments.

Report of the Research Briefing Panel on Scientific Frontiers and the Superconducting Super Collider

Research Briefing Panel on Scientific Frontiers and the Superconducting Super Collider

Stanley Wojcicki (*Chairman*), Stanford University, Palo Alto, Calif.
Barry Barish, California Institute of Technology, Pasadena
Kenneth Fowler, Lawrence Livermore Laboratory, Livermore, Calif.
Barry H. Honig, Columbia University, New York, N.Y.
Arthur K. Kerman, Massachusetts Institute of Technology, Cambridge
C. Kumar N. Patel, AT&T Bell Laboratories, Murray Hill, N.J.
Chris Quigg, Fermi National Accelerator Laboratory, Batavia, Ill.
David Schramm, University of Chicago, Ill.
Roy F. Schwitters, Harvard University, Cambridge, Mass.
Sam B. Treiman, Princeton University, Princeton, N.J.
Steve Weinberg, University of Texas, Austin
Frank Wilczek, University of California, Santa Barbara
Harry Woolf, Institute for Advanced Study, Princeton, N.J.

Staff

Donald C. Shapero, *Staff Director*, Board on Physics and Astronomy
Robert L. Riemer, *Rapporteur*, Board on Physics and Astronomy
Helene Patterson, *Staff Assistant*, Board on Physics and Astronomy
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

Report of the Research Briefing Panel on Scientific Frontiers and the Superconducting Super Collider

INTRODUCTION

Elementary-particle physics, the science of the ultimate constituents of matter and their interactions, has undergone a remarkable development during the past two decades. A host of experimental results made accessible by the present generation of particle accelerators and the accompanying rapid convergence of theoretical ideas has brought to the subject an unprecedented coherence. This activity has raised fresh possibilities and set new goals for understanding nature. The progress in particle physics has been more dramatic and more thoroughgoing than could have been imagined only a dozen years ago. Many of the deep issues then current have been addressed, and many of the opportunities then foreseen have been realized. This progress has brought us to an intellectual turning point comparable to the synthesis of classical physics that preceded the discovery of relativity and quantum mechanics in the late nineteenth century. We are thus compelled by these special developments in science to build an unprecedented facility to provide the experimental basis for further progress in our understanding of the most basic forces of nature.

Experimental investigation of some of the fundamental questions in elementary-particle physics requires energies higher than those provided by any accelerators now in operation or under construction anywhere in the world. For this reason, the U.S. elementary-particle physics community is now preparing a proposal for a very high energy superconducting proton-proton collider, the Superconducting Super Collider (SSC). This major new accelerator complex would be based on the accelerator principles and technology that were developed in connection with the construction of the Fermilab Tevatron and on extensive work on superconducting magnets in the United States over the past 20 years. The proposed super collider would have an effective energy range about 60 times higher than any collider now in operation and 20 times the energy of the Tevatron collider now nearing completion. The SSC is needed to answer some of the pressing questions in elementary-particle physics. In addition, the new energy range that would become accessible with the super collider represents new and uncharted territory, the exploration of which can be expected to bring about a number of important consequences.

In what follows, we summarize our current understanding in this field. We describe recent successes, both theoretical and experimental, and we review the questions raised by the standard model. The super collider complex is subsequently briefly described. The main part of this report treats specific examples of experiments that could be performed with the super collider that may elucidate current questions in particle physics and cosmology.

HISTORICAL BACKGROUND

Elementary-particle physics is the study of the basic nature of matter, energy, space, and time. Elementary-particle physicists seek to understand the fundamental constituents of matter and the forces that govern their behavior. In common with all physicists, they search for the unifying principles and physical laws that determine the nature of the material world around us.

The goals of particle physicists are a natural continuation of a perpetual human quest to understand the world. The belief that the universe is rational, that matter is composed of relatively few simple constituents, and that the laws governing them are fundamentally simple and comprehensible to the human mind has pervaded the history of science in the western world. These ideas were confirmed by the theoretical and experimental advances achieved by scientists in the nineteenth century. The periodic table of Mendeleev, a systematic organization of the information obtained up to that time on the nature of matter, is in some sense a summary of their achievement. The modern phase of physics began at about the same time with the study of atomic emission and absorption lines in the visible spectrum and the discovery of the electron.

The atom, the atomic nucleus, and the elementary particles of which they are composed are too small to see or study directly. Throughout this century, physicists have devised ever more sophisticated detection devices to observe the traces of these particles and their constituents, and they have created increasingly energetic beams of particles to probe more deeply into the structure of matter. Early examples are the use of x rays to probe the electronic structure of the atom and the use of radioactive sources to study the atomic nucleus.

A fundamental law of quantum mechanics is that increasing the energy of probing particles allows us to probe matter on a finer scale. The progression to ever-smaller distance scales achieved in this century from atoms to nuclei to nucleons (protons and neutrons) and finally today to quarks and leptons depended on the attainment of ever-higher energies. Thus particle physics today is the natural descendant of the atomic physics and nuclear physics of yesterday.

To achieve the ability to probe smaller distances, it was necessary to go beyond the naturally occurring sources of energetic projectiles from radioactive sources. Utilizing the fact that protons and electrons (the constituents of ordinary matter) are electrically charged and experience forces in the presence of electric and magnetic fields, scientists have devised increasingly sophisticated methods of accelerating them to high-energies. The colliding of high-energy particles and the analysis of collision products is at the heart of experimental particle physics. For this reason, the field is often called high-energy physics.

There are analogies between the situation in particle physics today and that in chemistry at the end of the nineteenth century. Elementary constituents (quarks and leptons), playing the role once played by the elements, have been identified, and the many regularities observed among them have been systematized. There was no fundamental understanding in the nineteenth century of the structure of the periodic table; similarly, we lack today an understanding of the laws that determine the observed pattern among the quarks and leptons. Revolutionary new ideas (quantum mechanics) were required to elucidate the reasons for the regularity of Mendeleev's table. Only time and further experimental data will tell us if equally dramatic new insights will be needed to explain present particle-physics puzzles. Before addressing those questions, it would be useful to summarize our present understanding in more detail.

THE REVOLUTION IN PARTICLE PHYSICS

Forty years ago, ordinary matter was thought to consist of protons, neutrons, and electrons. Experiments probed the structure of these particles and explored the forces that bind them together into nuclei and atoms. In the course of these experiments, physicists discovered more than a hundred new particles, called hadrons, which had many similarities to protons and neutrons. None of these particles seemed more elementary than any other, and there was little understanding of the mechanisms by which they interacted.

Since that time, a radically new and simple picture has emerged as a result of many crucial experimental discoveries and theoretical insights. It is now clear that the proton, neutron, and other hadrons are not elementary. Rather, they are composite systems made of much smaller particles called quarks, much as an atom is a composite system made up of electrons and a nucleus. The existence of five kinds of quark has been established, and initial experimental evidence for a sixth species has been reported.

Unlike our view of the proton and neutron, our view of the electron as an elementary constituent of matter that is structureless and indivisible has survived the revolution intact. However, we now know that there are six kinds of electronlike particles called leptons. Both quarks and leptons appear to be grouped in three families of two members each. According to our present understanding, then, ordinary matter is composed of quarks and leptons.

The world around us is made of just three of those fundamental particles—the two lightest quarks and the electron. The other elementary constituents of matter are created under extreme conditions such as may exist in collisions of energetic particles either in cosmic rays, accelerator beams, or the interiors of exotic stars. They were also presumably present in abundance some 15 billion years ago in the early stages of the creation of the universe.

Nature contrives enormous complexity of structure and dynamics from the six quarks and six leptons now thought to be the fundamental constituents of matter. Physics seeks to simplify the description of nature by finding the underlying causes of natural occurrences and by relating apparently distinct phenomena. The result of this effort has been to show that all natural processes may be understood as manifestations of a small number of fundamental interactions. For half a century, physicists have recognized four basic forces: gravitation, electromagnetism, the weak interaction responsible for certain radioactive decays, and the strong force that binds atomic nuclei. An important difference between quarks and leptons is that one of these four interactions, the strong force that binds quarks together to form hadrons, does not affect leptons. Both quarks and leptons are acted on by the three other fundamental forces.

Over the past two decades, great progress has been made in understanding the nature of the strong, weak, and electromagnetic forces. The weak and electromagnetic forces have been unified by a theory whose predictions have been verified by many inventive experiments, the culmination of which was the discovery of the *W* and *Z* particles in 1983. These carriers of the weak force are analogs of the photon, the carrier of the electromagnetic interaction, whose existence was postulated early in this century and established experimentally by the middle 1920s. In addition, there is indirect but persuasive evidence for particles called gluons, the carriers of the strong force. The strong, weak, and electromagnetic interactions are all described by similar mathematical theories called gauge theories. At present, the role played by the gravitational force in elementary-particle physics is unclear. The effect of gravity on the behavior of elementary particles is so small that it usually can be ignored.

The experimental measurements and discoveries that shaped the revolution in particle physics were made possible by harnessing new accelerator and detector technologies that permitted the exploration of new energy domains in novel and incisive ways. Accelerator advances included the exploitation of the strong-focusing principle in synchrotrons; the creation of intense high-energy beams of neutrinos; the invention of colliding-beam accelerators (colliders) in which counterrotating beams of high-energy particles collide head on; the development of bright sources of nearly monoenergetic antiprotons; and the introduction of large-scale, energy-efficient, high-field superconducting accelerator magnets. Among the advances in observational techniques were the utilization of the bubble chamber for the observation of reaction products and the parallel development of a series of ever more capable electronic detectors; the mastery of fast digital electronics for data acquisition and processing; the evolution of methods for managing and analyzing vast quantities of data; and the construction of large, complex detector systems exploiting the capabilities of a variety of individual devices. Each sortie into a new energy regime, each improvement in our ability to search for rare processes, and each increase in sensitivity for their detection has led to new insights and, often, to the discovery of unexpected and revealing phenomena.

With the identification of quarks and leptons as elementary particles and the emergence of gauge theories as descriptions of the fundamental interactions, we possess today a coherent point of view and a single language appropriate for the description of all subnuclear phenomena. This development has made particle physics a much more unified subject, and it has also helped us to perceive common interests with other specialties. One important by-product of recent developments in elementary-particle physics has been a recognition of the close connection between this field and the study of the early evolution of the universe from its beginning in a tremendously energetic primordial explosion called the big bang. Particle physics provides important insights into the processes and conditions that prevailed in the early universe. Deductions from the current state of the universe can, in turn, give us information about particle processes at energies that are too high to be produced in the laboratory—energies that existed only in the first instants after the primordial explosion.

WHAT WE WANT TO KNOW

Developments in elementary-particle physics during the past decade have brought us to a new level of understanding of physical laws. This new level of understanding is often called the standard model of elementary-particle physics. As usual, the attainment of a new level of understanding refocuses attention on old problems that have refused to go away and raises new questions that could not have been asked before. The quark model of hadrons and the gauge theories of the strong, weak, and electromagnetic interactions organize our present knowledge and provide a setting for going beyond what we now know. A useful analogy can be made between our present-day understanding of elementary-particle physics and the situation in atomic physics in the early twentieth century or in the beta-decay studies in the 1930s.

The Bohr model of the hydrogen atom was a radical idea that initiated a revolution not only in the field of atomic physics but also in our whole philosophical approach to physics. By incorporating into one theory a new concept of quantization of angular momentum together with the classical ideas of electrodynamics, Bohr succeeded in creating a formalism that was remarkably successful in explaining the observed spectral lines of hydrogen. But the model had obvious difficulties. Not only did it fail to predict the spectra of more complicated atoms, but it also left unresolved the fundamental question of why the electron does not fall into the center of the atom by radiating away all its energy. Nevertheless, on the basis of its partial success, scientists believed that the ultimately correct theory would almost certainly contain some of the ideas embodied in the Bohr theory.

A similar situation existed in the case of Fermi's weak interaction theory. The scheme was invented to explain the phenomena in the million-electron-volt (MeV) energy domain that characterize the spontaneously occurring beta-decay processes. Its extrapolation to higher energies was remarkably fruitful, since it successively provided the appropriate framework to describe muon decay and capture, strange-particle decays, and, finally, neutrino interactions at energies of billions of electron volts (GeV). The accuracies of the experimental agreements were startling because it was apparent that the theory was only a low-energy approximation that broke down when taken to the domain of hundreds of GeV. Because of its successes, however, one was confident that elements of the Fermi theory would eventually become part of a more complete description.

Particle physics may be in a similar situation today. Although the standard model provides a framework for describing elementary particles and their interactions, its success prompts us to seek a more comprehensive understanding. For example, we do not know what determines such basic properties of quarks and leptons as their masses. Nor do we understand fully the origin of the differences between the massless electromagnetic force carrier (the photon) and the massive carriers of the weak force (the W and Z particles). Existing methods for dealing with these questions involve the introduction of many unexplained numerical constants into the theory—a situation that many physicists find arbitrary and thus unsatisfying. Physicists are actively seeking more complete and fundamental answers to these questions.

Another set of questions goes beyond the existing synthesis. For example, how many kinds of quark and lepton are there? How are the quarks and leptons related, if they are related? How can the strong force be unified with the already unified electromagnetic and weak forces?

Then there are questions related to our overview of elementary-particle physics. Are the quarks and leptons really elementary? Are there yet other types of forces and elementary particles? Can gravitation be treated quantum mechanically as are the other forces, and can it be unified with them? More generally, will quantum mechanics continue to apply as we probe smaller and smaller distances? Do we understand the basic nature of space and time?

Given this list of questions, it is not surprising that there are many directions of theoretical speculation departing from the current paradigm. Many of these speculations imply important phenomena at energies that are at present beyond our reach. Although theoretical speculation and synthesis are valuable and necessary, particle physics cannot advance without new observations. In the recent past, crucial observations have come from a variety of sources, including experiments at accelerators and nuclear reactors, nonaccelerator experiments (cosmic-ray studies and the search for proton decay), and deductions from astrophysical measurements. All our current ideas, embodied in the standard model, point to one trillion electron volts (1 TeV is an energy equivalent to approximately 1,000 proton masses) as the energy scale on which new phenomena can be expected. But a diversity of experimental initiatives will have to be mounted in order to explore thoroughly the new regime of energy and distance. A detailed examination of a great variety of conjectured extensions of the standard model shows that the super collider, with the wealth of different experiments that it would support, is the instrument of choice for exploring this new domain. At the same time, these extensions of the standard model set the parameters for the new accelerator.

The super collider will not foreclose all the other investigations in the field of particle physics. The accelerators currently under construction (like the electron-positron colliders at the Stanford Linear Accelerator Center [SLAC] and the European Organization for Nuclear Research [CERN], and the Tevatron proton-antiproton collider at Fermi National Accelerator Laboratory) will provide detailed quantitative tests of the standard model and its relation to cosmology. More sensitive proton decay experiments, better neutrino mass measurements, improved searches for monopoles in cosmic rays, and more precise searches for forbidden decay processes are examples of other important lines of non-SSC investigations that will continue to be pursued. However, according to our present knowledge of elementary-particle physics, our physical intuition, and our past experience, most clues and information will come from experiments at the highest-energy accelerators.

DESCRIPTION OF THE SUPER COLLIDER

The super collider would have two counterrotating beams of protons guided by superconducting magnets. Each beam would be accelerated to 20 TeV and the two beams would be allowed to collide. Six different collision points around the circumference of the main ring are proposed in current designs. Sophisticated detectors would be installed at the interaction regions, or collision points. Many crucial experimental tests have already been framed, and conceptual designs of several complementary detectors have been carried out.

We must note that it is not the collisions of protons on protons as such that are of primary interest. As discussed previously, protons are composite systems formed of quarks and gluons, the latter providing the binding force that holds the proton together. It is these "hard" collisions of a constituent of one proton with a constituent of the other that provide us with information about the fundamental interactions. Very roughly, each constituent carries about one tenth of the proton's total energy and thus colliding beams of 20-TeV protons are required to produce numerous constituent-constituent collisions at a few TeV. The ease with which the relatively rare hard scatterings of constituents have been observed in proton-antiproton collisions at the CERN collider (with a beam energy of about 0.3 TeV) gives us confidence that important physics results can be derived from the complex collisions that would occur in a super collider.

The proposed accelerator scheme makes use first of an injector system consisting of a linear accelerator followed by two circular accelerators. This system accelerates protons to about 1 TeV, at which point they are injected into the main ring for the final acceleration phase. The diameter of the main ring will be somewhere between 30 and 50 km, depending on the strength of the magnetic field in the magnets chosen for the super collider. Several different magnet styles are currently under consideration. The eventual choice of the magnet style will be made with the objective of minimizing the construction and operating costs of the machine and maximizing its reliability.

All the magnets under consideration use the superconducting technology and cryogenic systems first successfully employed on a large scale at the recently completed Tevatron ring at the Fermi National Accelerator Laboratory. Superconducting magnets make the super collider feasible by significantly reducing the required power consumption and enabling one to achieve higher field strengths than are offered by the conventional room-temperature magnets.

Current super-collider designs use proven technology on a vast scale. The linear dimensions of the super collider would be at least 15 times larger than those of the Tevatron and 3 to 5 times larger than those of the collider (LEP) currently under construction by CERN. While the Tevatron requires about 1,000 superconducting magnets, the super collider would employ perhaps as many as 14,000. This very large scale presents extremely challenging problems in such areas as manufacturing techniques, quality control, reliability, civil engineering, instrumentation and controls, communications, and installation and repair logistics. The magnitude of these problems is rather new to accelerator science and technology but manageable by an appropriate extension of present skills and experience. Creative solution of these problems calls for a new partnership between the basic research community and industry, which may bring important advances in technology of practical value. The detectors provide another great challenge to the experimentalists. Only a tiny fraction of the 10^8 interactions occurring every second will be of interest. The techniques necessary to identify and record these interesting events in a small fraction of a second will advance the frontiers of electronics and computer technology.

SCIENTIFIC QUESTIONS FOR THE SUPER COLLIDER

A major accelerator facility is not constructed to carry out a single experiment or measurement but rather to make possible a great diversity of investigations over the accelerator's lifetime, which is measured in decades. The evolution of the experimental program is guided by results of early experiments, by improvements in detector and accelerator technology, and by theoretical insights. Because of this, we cannot describe in advance the full scope of the research program for a super collider. We indicate, however, some of the issues to be addressed in the first round of experiments. These are representative of the questions that form the basis for the scientific justification for constructing a super collider and of the opportunities that a super collider would present.

The Origin of Mass

Imagine a universe pervaded everywhere by a uniform magnetic field. In such a universe, the motion of charged particles would appear to be rather complicated, because the particles' motions would be influenced by the universal magnetic field as well as by specifically applied forces. Eventually, as physics developed, some genius would realize that a simple law of motion (namely Newton's) is really the basic one and that all the complicated spiralings observed in his universe are caused by a pervasive background field.

The current standard model of the weak interactions suggests that a similar situation is realized in our universe. The field involved is not a magnetic field but rather what is called a Higgs field. The equations describing the weak interactions would take a simpler and more symmetrical form if there were no background Higgs field. One simplification that would occur is that the masses of the *W* and *Z* bosons and of many other particles would vanish. This simplification evidently goes too far for the real world. The best we can do is to postulate that the basic model operates in a world pervaded by a background Higgs field that partially hides its full symmetry and simplicity. The standard model, with this standard assumption, accurately describes a wide variety of observations—including the existence and mass of the *W* and *Z* bosons recently found at CERN—but requires the existence of one or more scalar particles associated with this background field.

Certainly one main area that we can hope to clarify with the super collider is the exact nature of the Higgs field and its interactions with other matter. Why is the super collider likely to be an appropriate tool? To see the answer, consider the reason that the Higgs field pervades our universe. It must be because the total energy density is minimized by having this field present. To change the magnitude of the field significantly, or make manifest one of its particles, we must supply enough energy density to overcome the natural tendency of the field to revert to its normal universal background value. We estimate that the energies that would be available at the super collider are almost certainly sufficient to do so.

Our present theoretical understanding of Higgs fields is primitive. There must be at least one such field, but there may well be many, each associated with its own particle. The super collider will open a window on this now dimly perceived sector of elementary-particle theory and help us to understand the reason for the apparently chaotic pattern of elementary-particle masses, caused, according to the present theory, by the interactions with the Higgs field or fields.

Families

The recent discovery of the third family of quarks and leptons has sharpened the problem of understanding the replication of particles in families. Three families of elementary particles are known that have, within a very well-defined sense, identical strong, electromagnetic, and weak interactions but different masses. Why does nature repeat itself in this way? Are there still more families? Do Higgs particles come in families? Are there other particles that do not fit into the same repetitious pattern? There are many such questions; they will only be answered by experiments at higher energies than those that we can now attain.

One theoretical approach to understanding the existence of families postulates the existence of symmetries under which the different families are interchanged. Given such symmetries, the existence of one family necessarily implies the existence of others. A particularly appealing version of this idea involves combining the family symmetry and the gauge symmetries of the strong, weak, and electromagnetic interactions into one gigantic, all-encompassing symmetry, hidden by the presence of suitable background Higgs fields. Implementing such symmetries could lead one to expect a new class of family interactions mediated by heavier analogs of the *W* and *Z* bosons that might be created at super-collider energies.

Chirality

There is a peculiar asymmetry in the weak interactions of the observed quarks and leptons. Simply put, the *W* bosons prefer to couple to these particles if they appear to be spinning clockwise, as viewed by an observer that they are approaching. We say that the weak interactions mediated by *W* bosons couple to left-handed quarks and leptons. It turns out that they couple to right-handed antiparticles (e.g., antiquarks or positrons). Many people have speculated that this asymmetry arises by a mechanism analogous to the one that causes the asymmetry between electromagnetic and weak interactions to appear in the standard model. According to this view, there will be either *W'* bosons that couple preferentially to right-handed rather than left-handed quarks and leptons, or new sets of quarks and leptons with right-handed couplings to the familiar *W* boson. In many models, the symmetry between left and right is broken by the same Higgs fields that spoil the fundamental weak-interaction symmetry. In such models, the masses of the *W'* and *Z'* bosons involved, or of the right-handed quarks and leptons, cannot be much larger than the known mass of the *W* and *Z*—which would make them observable at super-collider energies.

Supersymmetry

Recently there have been many theoretical investigations of the physical consequences of a radically new kind of symmetry called supersymmetry that implies the existence of integral-spin replicas or partners of particles of half-integral spin and vice versa. If it is relevant to physics at all, supersymmetry must be a broken symmetry—for example, there is definitely no particle having the mass and charge of the spin-1/2 electron that has integral spin. Such a particle would have been observed long ago if it existed.

There may be significant advantages to the view that supersymmetry is not too badly broken, so that many supersymmetry partners of known particles could be produced at super-collider energies. These advantages include resolution of the mathematical inconsistencies arising in the calculation of many physical quantities—specifically the masses of the Higgs particles.

The conjectured supersymmetry would lead to a Higgs boson mass less than 1 TeV/c^2 and yield supersymmetric particles with masses also less than about 1 TeV/c^2. Up to the present, there is no conclusive experimental evidence for these superpartners, even though there have been attempts to interpret some unusual events observed at CERN with large amounts of energy radiated in an invisible form in terms of supersymmetric models.

Significant progress has been made recently in building unified field theories that include gravity. These ambitious theories, which incorporate supersymmetry in a fundamental way, involve objects called superstrings; they will be confirmed, significantly constrained, or ruled out by experiments that elucidate the possible role of supersymmetry at super-collider energies.

Dynamical Symmetry Breaking

A second possible solution to the Higgs problem assumes that the Higgs boson is not an elementary particle at all but a composite object made of elementary constituents analogous to quarks and leptons. Although they would resemble the usual quarks and leptons, these new constituents would be subject to a new kind of strong interaction that would confine them within about 10^{-17} cm. Such new forces could yield new phenomena as rich and diverse as the conventional strong interactions, but on an energy scale a thousand times greater—around 1 TeV. The new phenomena would include a rich spectrum of new bound states akin to the spectrum of known hadrons. Again, there is no evidence yet for these new particles.

Compositeness

Violent collisions among quarks also provide a window on the possible internal structure of quarks. Some physicists feel that the known quarks and leptons are too numerous to be the ultimate elementary particles. If quarks are themselves composite, it should be possible to excite their internal structure in violent collisions. One sign of this kind of internal excitation would be spectacular multijet events quite unlike those anticipated in the standard model. The absence of such events and the agreement of the observed quark-quark scattering cross section with the standard-model predictions imply upper limits on the size of quarks. It will be possible with the super collider to look for quark substructure down to a distance of about 10^{-18} cm.

In summary, both general arguments and specific conjectures for resolutions of the Higgs problem imply 1 TeV as an energy scale on which new phenomena crucial to our understanding of the fundamental interactions must occur. The origin of electroweak symmetry breaking is only one of the important issues that define the frontier of elementary-particle physics. However, because of its immediate and fundamental significance it must guide our planning for future facilities.

Cosmology and the Super Collider

Over the past few years, cosmology and particle physics have become increasingly interwoven. To understand the physics that took place in the high-temperature, high-density early universe, one is forced to look at the physics of elementary particles. Similarly, the unified theories of elementary-particle physics have striking consequences at extremely high temperatures and energies. The only laboratory available to check these extrapolations of unified theories is the first instants after the big bang when extraordinarily high temperatures and densities were reached. In some ways, the two fields have become symbiotic. The super collider will be operating at energies far beyond those previously achievable in a laboratory and will simulate the conditions that prevailed when the temperature of the universe was about 10^{17} K. These conditions occurred about 10^{-15} second after the primordial explosion. Direct observations by optical telescopes are limited to events that occurred roughly 300,000 years after the big bang because the universe was opaque to photons at earlier times.

To reconstruct what occurred in the early universe, it is necessary to know the nature of fundamental interactions at high energies and the complete spectrum of elementary particles. In particular, the relics left over from those early times are of fundamental importance to cosmology. Any long-lived particle produced in the primordial explosion would survive and be an ingredient in the present-day universe.

One of the major issues in cosmology is to find the dark matter of the universe. Studies of the motion of stars within galaxies and of galaxies within clusters have established that these systems must contain a great deal of matter in addition to what is visible in the stars. This nonluminous matter may in fact account for the bulk of the mass in the universe. The properties that we impute to the dark matter depend on the character of the small density fluctuations in the early universe that grew into the galaxies and clusters observed today. According to current ideas on galaxy formation, the dark matter may be quite different from the ordinary (baryonic) matter of which we are made. Particle physics yields a mechanism for generating the primordial density fluctuations and provides candidates for the dark matter as well. Experimentation at the super collider will allow broad searches for new particles that may play the role of the dark matter.

In addition to the possibility of illuminating the question of dark matter of the universe, the super collider will clarify the structure and symmetry of the fundamental interactions and allow us to extrapolate with greater confidence back to early times. One of the most interesting recent developments in cosmology has been the suggestion that the homogeneity and isotropy of the universe were established in an early phase transition. According to this scenario, the universe began in a highly symmetric phase in which all the fundamental interactions were equivalent and evolved to the present low-symmetry phase in which different forces have different manifestations.

Another problem in astrophysics concerns unusual events produced in the infrequent collisions of very high energy cosmic rays with the Earth's atmosphere. Some of these events indicate unexpected phenomena occurring in the range of energies that can be explored systematically by the super collider. Such an exploration would make possible unambiguous interpretation of the hints provided by ultra high energy cosmic-ray events.

CONCLUSION

The advances of the past decade have brought us tantalizingly close to a profound new understanding of the fundamental constituents of nature and their interactions. The standard model based on quarks and leptons organizes our current knowledge and defines the frontier of particle physics at constituent energies of about 1 TeV and the frontier of cosmology at times of about 10^{-15} second. There we await new discoveries about the unification of the forces of nature, the patterns of the fundamental constituents of matter, and the origin of the universe.

Report of the Research Briefing Panel on Computer Vision and Pattern Recognition

Research Briefing Panel on Computer Vision and Pattern Recognition

Azriel Rosenfeld (*Chairman*), University of Maryland, College Park
J. K. Aggarwal, University of Texas, Austin
William L. Alford, Defense Mapping Agency/STT, Washington, D.C.
Norman Caplan, National Science Foundation, Washington, D.C.
Su-shing Chen, National Science Foundation, Washington, D.C.
Chao K. Chow, IBM Corporation, Yorktown Heights, N.Y.
Thomas M. Cover, Stanford University, Stanford, Calif.
Martin Fischler, SRI International, Menlo Park, Calif.
King-sun Fu, Purdue University, West Lafayette, Ind.
Robert Haralick, Machine Vision International, Ann Arbor, Mich.
Richard Heydorn, Johnson Space Flight Center, Houston, Tex.
Takeo Kanade, Carnegie-Mellon University, Pittsburgh, Pa.
Ernest Kent, National Bureau of Standards, Gaithersburg, Md.
Robert D. Leighty, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Va.
George Nagy, University of Nebraska, Lincoln, Nebr.
Theodosios Pavlidis, AT&T Bell Laboratories, Murray Hill, N.J.

Staff

Gary E. Clark, *Staff Director*, Computer Science and Technology Board
Allan R. Hoffman, *Executive Director*, Committee on Science, Engineering, and Public Policy

Report of the Research Briefing Panel on Computer Vision and Pattern Recognition

1. INTRODUCTION

Computer vision is the process of deriving information about a scene by computer analysis of images of the scene. The images are obtained by a sensor, such as a television camera, and are converted to digital form for processing by a computer. The goal is usually the recognition of objects that appear in the scene. The analysis makes use, in part, of techniques from fields such as image processing, pattern recognition or classification, and artificial intelligence.

The field of computer vision is over 30 years old. As far back as the 1950s work was being done on such tasks as reading printed characters, counting blood cells, and recognizing military targets by analyzing digitized images. By the 1960s character reading had become a commercially successful application, and by the 1970s systems for medical applications such as blood counting had become commercially available. Research on military applications continued to be supported throughout this period, leading to successful demonstrations in areas such as tactical target detection and recognition and to practical use in terminal missile guidance. The late 1960s saw the beginnings of research on robot vision systems as well as on other industrial applications such as product inspection. By the late 1970s such applications began to reach the factory floor, and today there are more than 100 companies marketing industrial computer vision systems and devices.

As these examples indicate, computer vision has had many practical successes, but there are still limits on its capabilities. Some of these limits are due to inadequate computer power, but others are of a more fundamental nature. It has been estimated that adequate computer analysis of some types of complex, real-word scenes will require on the order of billions of computer operations per image. The computer power required to do this at television rates (30 images per second) is several orders of magnitude beyond what is available today at reasonable cost. The algorithms that will be needed for such complex analyses are not yet well understood—in part, because of the high computational cost of experimenting with such algorithms. More fundamentally, there is as yet insufficient theoretical basis for designing algorithms that can handle complex scenes; not enough is known about the mathematical description or modeling of classes of scenes.

The recent rapid advances in computer

hardware technology have made it possible to design special-purpose, massively parallel computer architectures suitable for the fast implementation of vision algorithms. The advent of VLSI (Very Large Scale Integration of digital circuits) and related hardware developments will greatly alleviate today's computer power limitations on vision systems. Efforts now need to be undertaken to alleviate the algorithmic and theoretical limitations. A significant research program is needed to stimulate advances in vision algorithms and in the theoretical foundations that underlie them. As successful algorithms are developed, it will be possible to design architectures that are matched to their computational requirements. Conversely, the availability of new architectures will make it possible to conduct significant experiments with complex vision algorithms and to learn more about their properties and about how to design them. The opportunities for major advances in computer vision over the next decade are exciting and challenging.

Section 2 of this report describes representative computer vision techniques designed for tasks of various degrees of complexity, and indicates the limitations on these techniques. Section 3 summarizes the state of the art in several major application areas, including remote sensing, cartography, reconnaissance, navigation, robotics, industrial inspection, and document understanding, and lists what appear to be the necessary conditions for successful applications of computer vision using today's technology. Section 4 outlines a set of research and development needs aimed at expanding the range of feasible applications. Finally, Section 5 discusses possible strategies for maximizing the payoff resulting from such a research program.

2. COMPUTER VISION TECHNIQUES

Before describing some typical computer vision techniques, we introduce some basic terminology. A computer vision system deals with images obtained from a *sensor* that views or scans the scene. In order to process these images by digital computer, they must first be converted into digital form, i.e., into arrays of numbers. Such an array is called a *digital image*, and its elements are called *pixels* (short for "picture elements"). The value of a pixel represents the brightness of the image in the neighborhood of a given point. This brightness, in turn, results from the radiation received by the sensor from the scene along a given direction. The number of pixels should be sufficient to adequately represent the resolution of the sensor. Conventional television images are usually digitized into arrays of about 500 by 500 pixels.

2.1 Pixel Classification

In remote sensing of the earth's surface or the atmosphere from satellite altitudes, a single pixel represents a significant area, usually tens of meters in diameter, on the earth's surface. Typically, the sensor produces a number of images using different spectral bands in the visible and near infrared. When these images are digitized, we thus have a set of values associated with each pixel.

Such multispectral images have been extensively used for terrain or crop classification. The standard approach is to assign each pixel to a class on the basis of its set of spectral measurements, using statistical decision techniques.

The accuracy with which this type of classification can be carried out depends on how accurately we model the problem domain. If we simply characterize the classes by their spectral measurements, we find that these measurements are highly variable. As a result, classification based on statistical decision methods will have a high inherent error rate. The error rate is further increased by the fact that the ground area corresponding to a pixel may contain a mixture of classes. Improved results can be obtained by using models that incorporate physical information about the scene.

2.2 Two-Dimensional Object Recognition

In most applications of computer vision, the objects to be recognized are much bigger than a single pixel. We consider first the situation where the scene is essentially two-dimensional, for example, where the objects are characters on a document, lines on a drawing, or connectors on a circuit board. In this situation, any objects present in the scene should be fully visible on the image.

If the objects to be recognized are exactly known—for example, if they are characters from a known typeface—a straightforward "template matching" process can be used to recognize them. The main limitations on recognition accuracy are the precision with which the object shapes are known and the fidelity with which the image represents the scene; accuracy decreases if the image is degraded (has low contrast, is noisy or blurred) or if the objects themselves are imperfect (e.g., badly printed characters). We can optimize system performance if we have good models for both the object and the degradation processes.

In most situations, object shapes are not known exactly, and recognition accuracy is limited not only by image degradations but also by how well we can define (i.e., model) the shapes belonging to a given object class. A classical approach is to characterize the objects by a set of geometrical measurements or features (e.g., area, perimeter, number of holes) and to use statistical decision methods. If the classes to be recognized are sufficiently dissimilar, this approach can yield acceptable levels of recognition. Of course, its success depends on our ability to extract ("segment") the objects from their background so that the features can be measured correctly. Successful segmentation depends in turn on being able to characterize (i.e., model) how the objects differ from the background (e.g., characters are usually darker than the surface on which they appear).

For complex objects, the feature-based approach is usually not adequate to capture the subtle differences that distinguish the classes. Rather, it becomes necessary to model the objects in greater detail, e.g., as composed of parts that have given features and that are arranged in given configurations. (For example, we can model characters as composed of strokes that have specified shapes and are combined in particular ways.) To apply this "structural" approach, it is necessary to extract the parts from the image, measure their features, and determine what relationships hold among them. This information can be represented by a graph structure, and recognition can be achieved by matching this structure with stored graph structures representing the models for the given classes of objects. (Note that graph matching may involve searching through a very large space of possible matches.) This approach is much more flexible than feature-based techniques, but it may require models that are complex and difficult to specify adequately.

2.3 Three-Dimensional Object Recognition

Computer vision becomes much more difficult when it is necessary to deal with truly three-dimensional scenes—for example, when a robot has to recognize mechanical parts in a pile, or plan a path across a cluttered factory floor. Two factors make such three-dimensional tasks more difficult:

a. The image shows only a two-dimensional projection of the scene; the visible surface points in the scene lie at different distances from the sensor, but the position of a point in the image tells us only along what direction in the scene the corresponding surface point lies, not how far away it is.

b. Even if we knew the distances, we would only have information about the surfaces that are visible from the sensor; we cannot see the hidden sides of objects, or the parts of objects that are hidden by other objects.

A class of techniques has been developed over the past 15 years to deal with the first problem; these techniques are known collectively as "shape from . . ." methods. They infer information about the three-dimensional shapes of the visible surfaces in the scene from a variety of clues present in the image: brightness variations ("shading"), changes in the sizes and spacings of local patterns ("texture"), etc., by making use of assumptions about (i.e., models for) the shapes and surface properties of objects and the sources of illumination. The ultimate goal of these techniques is to determine the illumination, reflectivity, and spatial orientation of each visible surface point; these are the factors that give rise to the brightness of the corresponding image point. In effect, we are attempting to solve the "inverse optics" problem; if we know the values of these factors over the scene, it is easy to compute the image brightness, but the reverse is much harder. If we can estimate the values of these factors at each pixel, we now have a set of values, rather than a single value (brightness) associated with the pixel. In other words, we have replaced the original image by a set of images. These are known as "intrinsic" images, since each of them represents an intrinsic physical property computed at each visible point of the scene. The process of estimating the intrinsic image information is known as "recovery" (short for "recovery of intrinsic scene characteristics from an image").

Extraction of objects from an image of a three-dimensional scene becomes much easier if intrinsic image information is available. An object and its background may not be easily distinguishable on the basis of their brightnesses in the image; even if they have different reflectivities, variations in illumination and surface orientation may cause their brightness ranges to overlap. However, if we know the illumination and orientation, we can correct for their effects.

Even if we have intrinsic image information, recognition of three-dimensional objects is still difficult, since an object may be partially hidden by other objects. Even if it is not hidden, it may occur in the scene in many different orientations, so that its projection in the image may have many different shapes. One approach to recognizing objects in arbitrary orientations is to use a large set of two-dimensional shape models, one for "every" orientation. A more flexible approach makes use of a set of three-dimensional models for object parts and is based on an "inverse projective geometry" technique. Given a region in the image, not all of the possible three-dimensional object parts can project into it, and for those that can, not all orientations are possible. Suppose we find a set of regions all of which could be the projections of parts of the same three-dimensional object, all oriented consistently; then we have strong evidence for the recognition of that object.

2.4 Other Techniques

In this section we briefly mention some aspects of computer vision that were omitted from the previous sections for the sake of simplicity.

Local feature detection A region in an image may have a wide range of brightnesses, but it may still be distinguishable from its background because there is a sharp brightness discontinuity along its border. Many techniques for detecting such brightness "edges" in images have been developed, based on models for various types of physical discontinuities in the scene. There are also techniques for detecting special types of local image features involving brightness discontinuities—for example, thin lines or curves that contrast locally with their backgrounds.

Texture analysis A region in an image is sometimes distinguishable from its background because it has a characteristic pattern of brightnesses or local features. Many techniques have been developed for modeling and discriminating among such patterns, or "visual textures," and using them as a basis for segmenting an image. As mentioned earlier, texture variations across a region in an image can also be a useful clue to the three-dimensional orientation of the corresponding surface in the scene.

Stereo If two images of a scene taken from different (known) positions are available, and we can identify corresponding points (arising from the same scene point) in the two images, we can determine the three-dimensional location of the scene point by triangulation. The images are called a stereopair, and the triangulation process is called stereoplotting. The difficult problems are (1) reliably identifying corresponding pairs of points and (2) interpolating plausible three-dimensional surfaces based on a sparse set of point correspondences.

Motion Given a sequence of images of a scene taken at closely spaced times from a moving sensor, by comparing successive images we can derive information about the three-dimensional structure of the scene. Many techniques have been developed for measuring the "optical flow" of intensities in a time sequence of images and for inferring scene structure from the flow field.

Controlled illumination Control of illumination can make computer vision tasks easier in a variety of ways. It can be used to improve the contrast between an object and its background, e.g., by silhouetting the object. Three-dimensional information about a surface can be unambiguously derived by comparing the intensity variations across the surface when it is illuminated from several known directions; this method is called "photometric stereo." Patterned illumination can also provide three-dimensional surface information with the aid of triangulation techniques; this is known as "structured light" range measurement.

3. APPLICATION AREAS

3.1 Remote Sensing

Pixel classification based on spectral measurements, as described in Section 2.1, has been extensively used for crop classification and crop acreage estimation. The discussion here deals primarily with measurements obtained by satellite-borne multispectral scanners, but there are other applications that make use of color aerial photography. An important extension is to use several multispectral images taken at different times during the growing season.

Characterizing crop classes by their spectral measurements is not entirely adequate, since these measurements are sensitive to variations in soil background and planting data, as well as to atmospheric effects. Typical accuracies obtained, e.g., for discriminating corn from soybeans or winter grains from other crops, are only about 70 percent. Better results, with accuracies of about 85 percent, are obtained by introducing models for the growth cycle of the crops. We now represent the data by a set of estimated model parameters at each pixel; this transformation of the data often makes the classification problem easier by emphasizing parameters that best discriminate crops and ignoring other factors.

Further improvement should be achievable by introducing more detailed models that take into account the physical structure of the ground cover. One can model the scattering of radiation from vegetation in terms of parameters such as leaf angle distribution, leaf transmittance, and leaf reflectance, as well as the geometric placement of the plants. Of course, such models are quite complex, and extensive studies will be needed to define them quantitatively for a given class of crop scenes. The observed measurements often result from a mixture of several models, and it will also be necessary to develop general methods of handling such mixtures.

3.2 Cartography and Reconnaissance

Map makers use stereopairs of aerial photographs as a primary source of information about the earth's surface, including both measurement of terrain height and recognition of surface features. Systems proposed for operational use may incorporate some degree of automated stereoplotting, but are still highly dependent on interaction with human operators. By the 1990s, operational systems may incorporate computer vision capabilities for automated extraction and recognition of various feature types, such as hydrological features, vegetation, roads, and certain types of structures. In relatively flat terrain viewed from high above, feature recognition is essentially a two-dimensional problem, and can usually be handled by the methods described in Section 2.2.

Airborne reconnaissance is conducted in two modes: searching and monitoring. In the searching mode, targets can occur anywhere and the task is to detect and recognize them. The monitoring mode involves periodic examination of a known area to check on its operational status. In both modes, today's systems can provide computer assistance to a human operator. Selected types of targets, e.g., vehicles in infrared images, can be searched for, detected, and their positions signaled to the operator (which is known as "target cueing"). Systems now in the research and development stage are expected to achieve reliable target recognition capabilities through the use of contextual information.

3.3. Navigation

Terminal homing is the process of guiding a missile to a target by correcting its path with the aid of computer vision techniques. This generally involves matching a stored template of the target with a sensor image. Such techniques are used in the Pershing missile system. A proposed improvement will make use of a set of templates corresponding to different target ranges.

Mid-course guidance is the process of determining the location of a missile in order to correct its course. Present cruise missiles do this by matching a stored pattern of terrain elevations with radar altimeter data. A more robust approach would match stored templates with a high-resolution sensor image; this should be feasible now or in the near future.

There has been recent interest in the development of an autonomous land vehicle capable of traversing a road network, and ultimately of navigating across various types of terrain. Such a vehicle will make extensive use of computer vision capabilities for road detection and following, obstacle detection and avoidance, and detection of landmarks for navigation purposes. The Defense Advanced Research Projects Agency (DARPA) has initiated work on such a system under its Strategic Computing Program. Autonomous land vehicles, or mobile robots, will have many nonmilitary uses in areas such as construction, shipbuilding, and agriculture.

3.4 Industrial Inspection and Robot Vision

Most of today's industrial computer vision systems are designed to perform inspection tasks. Such tasks can be reliably automated, using today's technology, if the following conditions hold:

a. The object to be inspected has a regular or known shape, having sharply defined, high-contrast features;
b. Objects do not overlap, and have known orientations;
c. The lighting is controlled and uniform;
d. Defects are large and clear cut.

An example of an application area for which many working systems have been developed is printed circuit inspection.

Another important industrial application of computer vision is in robot vision systems. Most systems in current commercial use employ very simple techniques in highly constrained situations. A typical application is seam tracking in robot welding, using structured light triangulation techniques. Other systems operate on high-contrast images by first segmenting them into objects and background; this enables the system to deal with one-bit pixel arrays in which 1's are object pixels and 0's are background pixels. If the segmentation can be done reliably, perhaps with the aid of controlled lighting, such "binary" vision systems can operate at high speeds. One application for such systems is the lifting of objects from a conveyor belt, where the objects can have only a few stable positions. Picking objects out of a bin is a much more difficult task, since the objects can be in arbitrary three-dimensional orientations and the lighting inside a bin cannot be completely controlled.

Robot vision will find its greatest use in situations where the timing, placement, orientation, and identity of arriving objects are not completely constrained. Ultimately, robots will need to recognize and locate specific objects in piles of mixed objects. Techniques have been developed that can handle such tasks, but they are still computationally costly and slow.

3.5. Document Understanding

Document understanding started with typewritten optical character recognition and has been expanded to include recognition of typeset copy, engineering drawings, maps, and other graphical information. While there is no "universal" machine for reading arbitrary documents, there have been many partial successes.

The main factor influencing success in this area has been control of the input. Chronologically, the problems solved were specially shaped numerals on checks; bar codes; monospaced, single-font readers; readers for a limited number of fonts; and reading of carefully formed hand-printed letters and numerals within boxes.

A second factor has been the use of prior knowledge or context. For example, in postal sorting of outgoing mail the postal code is compared with the city and state and a decision is made only if both agree. Similarly, in scanning line drawings of logic circuits, the graphic input is checked against circuit constraints.

Areas ripe for near-term development include multifont typeset text, typeset text interspersed with graphics, and graphics labeled with typeset text. An important point in general document processing is that it is not sufficient to classify the symbols. It is also essential to represent the geometric relations between them, since these too convey meaning.

3.6. Characteristics of Successful Applications

In each of the domains discussed above, many tasks can be successfully carried out using today's computer vision technology, but many other tasks still present research challenges. The following are some of the criteria that characterize today's successful applications:

a. Computer vision is easier in a restricted scene domain, in which the class of scenes can be modeled with sufficient accuracy. The domain can sometimes be simplified by controlling the scene environment, for example, by controlling the lighting.

b. Two-dimensional scenes are easier to deal with than three-dimensional scenes.

c. There is a trade-off between the time available to carry out a task and the computer power required. If a task must be performed under very tight time constraints, its implementation may require impractical amounts of computation.

An example of a task that is extremely difficult in terms of all these criteria is autonomous vehicle navigation. Here the scene domain is potentially unrestricted; the environment is uncontrolled; the scene is strongly three-dimensional; and the vehicle must navigate in real time.

In order to extend computer vision capabilities to domains that do not satisfy the above criteria, a substantial research and development effort will be necessary, as discussed in the next section.

4. RESEARCH AND DEVELOPMENT NEEDS

4.1. Sensors

Most of the discussion in previous sections has assumed that the sensor is some type of television camera, but many other types of sensors have been used in computer vision systems. The airborne and satellite-borne scanners used in remote sensing detect radiation in a set of spectral bands in the visible and near infrared. Thermal infrared imaging sensors that provide temperature information are often used in target detection applications. Sensors that operate in other spectral bands can also be used; for example, extensive research has been done on the application of computer vision techniques to x-ray images, both for medical purposes (for example, analysis of heart shape or detection of tumors) and for industrial inspection. One can also work with sensors that make use of nonelectromagnetic energy; for example, ultrasonic arrays are used for obstacle detection in various applications.

An important class of sensors provides three-dimensional information about a scene, rather than just intensity information. Radar sensors directly measure the range to every visible scene point, as well as providing information about the dielectric properties of surfaces. Conventional radars use microwave radiation, but systems using visible light have also been developed. The use of such sensors greatly reduces the difficulty of handling three-dimensional scenes, but it still does not provide information about hidden parts of objects. Under the proper circumstances, complete three-dimensional information about the scene can be obtained using tomographic reconstruction techniques, which yield a three-dimensional array of "pixels" representing "density" information at each point of a three-dimensional volume. Some work has been done on extending computer vision techniques to handle such three-dimensional arrays of data.

As we have seen, successful design of a computer vision system depends on adequately modeling the scene, the objects present in it, and the process by which the images are obtained. Most existing computer vision techniques were designed for images obtained by optical or near-infrared sensors. For sensors that make use of other types of radiation, such as x rays or ultrasound, the imaging process is quite different, and the computer vision techniques used should be based on different models. These remarks apply even more strongly to sensors that provide direct range information or complete three-dimensional information. As such sensors begin to be used in computer vision systems, new techniques will have to be developed based on the appropriate models.

In many situations, more than one sensor will be needed in order to carry out a given task, since different sensors will provide different types of information. This leads to the need to develop techniques for integration or "fusion" of information from different sensors. Registering the data obtained by different types of sensors is a nontrivial task, since the sensors generally have different resolutions and often have different imaging geometries.

Even when conventional sensors are used, the choice of sensor parameters may significantly affect the difficulty of computer vision tasks. There is an important engineering trade-off between sensor cost and computational cost. In some applications, for example in the document processing area, it has been shown that a moderate increase in sensor resolution and linearity made it possible to avoid a massive increase in computation.

4.2. Computers

Digital images contain voluminous amounts of data—typically, hundreds of thousands of pixels. Typical computer vision tasks may require hundreds of computer operations per pixel, for a total of tens or hundreds of millions of operations per image. If images need to be processed at television rates (30 per second), conventional computers are not fast enough.

The solution to this problem lies in the use of massive multiprocessing. The match between the computational needs of computer vision and the emerging computational capabilities of multiprocessing systems appears to be excellent. Computer vision tasks often involve the performance of identical computations, or sequences of computations, at every point in an image. Multiprocessing makes it possible to carry out such tasks in parallel by assigning different regions of the image to different processors. These processors can operate relatively independently; for many operations, only local communication between neighboring processors is needed. Long sequences of computations can be carried out by "pipelined" multiprocessors, with each processor operating on the output produced by the preceding one.

Major research and development efforts are under way worldwide to build multiprocessor systems for computer vision and image processing. Dozens of such systems have been designed and built in the United States, Japan, and Western Europe. Experiments with these systems will provide us with useful information about the types of multiprocessor architectures that will be most useful for computer vision.

The recent advances in VLSI make it possible to produce powerful special-purpose processors in quantity. A wide variety of different processor architectures can be used to speed up computer vision processes. It is now possible to design and fabricate chips that implement specific image processing, pattern recognition, and computer vision techniques, for example, template matching. By combining such chips very fast vision computers could be produced.

Since computer vision systems usually employ optical sensors, some types of initial processing of the sensor data can be done at very high speeds, in parallel, using optical filtering techniques. Optical and hybrid optical-digital methods can be used to perform a variety of useful operations on images.

4.3 Models

We have emphasized earlier that to achieve high levels of performance, computer vision systems should make use of good models for the imaging process and for the classes of scenes to be analyzed. Many types of scene models are highly specialized (e.g., models for crop/leaf canopies), while others are of a more general nature (e.g., Markov models for textures, fractal models for surfaces, reflectivity models). In some areas, good modeling techniques do not as yet exist; better methods are needed for modeling classes of complex object shapes (in two or three dimensions) or spatial relations between objects. More intensive work is needed on developing and improving models for commonly analyzed classes of scenes.

Models can provide a theoretical basis for designing computer vision systems. They make it possible to design optimal techniques for such problems as image segmentation, local feature detection, texture discrimination or recovery of intrinsic scene information, optimal sets of features for two- or three-dimensional shape recognition, and so on, and to predict the performance of these techniques. It should be pointed out that many of these problems are underconstrained or "ill-posed," and can be solved only if some type of model for the scene domain is specified.

In nearly all existing computer vision systems, models are used implicitly as a basis for designing the analysis techniques used by the system. An important current research area is the design of "expert" computer vision systems in which the models (i.e., the system's knowledge about the scene, the sensor, and other factors) is represented explicitly. By making this information explicit, it becomes possible to design the system to evaluate its own performance, determine the possible causes of errors, and take appropriate action. The addition of such capabilities to computer vision systems would lead to substantial improvements in their performance.

4.4. Computational and Algorithmic Tools

Computer vision systems make use of many computational techniques and tools that are also used in other areas of pattern recognition, artificial intelligence, and computer science. Improvements in these tools would be of great benefit to computer vision research. The following are some examples of techniques that are of particular potential importance:

a. Pattern matching techniques, including matching of images and of graph structures derived from images;

b. Methods of combining evidence obtained from multiple sources and of resolving conflicts;

c. Methods of effectively using context in decision making and recognition;

d. Techniques for controlling potential "combinatorial explosions" in the size of search spaces;

e. "Indexing" techniques for rapidly selecting, from a large data base, the models that are relevant to the current task;

f. Methods of partitioning problems into coherent components that can be analyzed independently;

g. Methods of communicating information between processes that are operating in parallel;

h. Knowledge acquisition techniques for building and organizing models or knowledge bases for given problems;

i. Techniques for effective human-machine communication to discuss iconic and spatial data.

4.5. Theoretical Foundations

Many of the specific processes used in vision systems have well-developed theories; for example, the techniques for recovering intrinsic scene properties from an image have a well-understood mathematical basis. In many parts of the field, however, ad hoc techniques are still extensively used. The development of theoretical foundations in other parts of the field would provide a basis for designing optimal algorithms, conducting controlled experiments, determining their statistical validity, and evaluating their results.

A more challenging task would be that of developing a general theory of the overall computer vision (or pattern recognition) process. A possible basis for such a theory lies in the concept of complexity. Kolmogorov has defined the complexity of an object as the length of the shortest computer program for describing that object. This notion of complexity turns out to be independent of the particular computational model that is used; it is a universal notion. The goal of image description is to derive just such concise descriptions for a given image. More precisely, given a language in which such descriptions can be formulated—i.e., a language for describing the given class of scenes—the goal is to find "good" or "sufficient" descriptions, in that language, for a particular image. This formulation is conceptually important even though, in real-world cases, it may be computationally impractical to realize. (The computation time required to find the shortest description of an object is, in general, unbounded.) Nevertheless, it would be desirable to assess the current ad hoc techniques in relation to the domain of theoretically achievable techniques—in particular, to assess how far away they are from the theoretical optimum.

In general, the field of computer vision can benefit from research in other related areas, including artificial intelligence, automatic inference, computational complexity, nonparametric modeling, and statistical pattern recognition. Computer vision researchers can also benefit from reviewing what is known about biological visual systems, since these systems constitute an existence proof that complex visual tasks can be carried out rapidly and reliably. The brain's ability to recognize objects in complex scenes in a fraction of a second demonstrates the power of massively parallel computation. Conversely, the design of special computer architectures for real-time computer vision may provide clues about the structure of the brain, and computer vision algorithms may suggest models for biological visual processes.

5. RESEARCH STRATEGIES

In this section we discuss some alternative strategies for supporting the research that will be needed to achieve substantial advances in computer vision. Such support will have to come primarily from federal sources, since it represents a long-term investment. Computer vision activities in the private sector have been confined almost exclusively to short-term product developments having predictable payoffs and involving relatively simple applications, in areas such as character recognition and two-dimensional industrial inspection.

The conventional approach to the support of basic research in a given area has been to fund a number of relatively small individual grants addressing specific aspects of the subject. This is the approach followed by the National Science Foundation, for example. As applied to computer vision, it has produced, and will continue to produce, important advances in the underlying theory, paradigms, principles, and techniques.

DARPA has adopted a different strategy in connection with its Strategic Computing Program. It has defined a number of major application goals, the achievement of which will require substantial advances on many fronts. In the area of computer vision, the application is embodied in the Autonomous Land Vehicle project, which will accomplish a series of milestone demonstrations involving road following, obstacle avoidance, and route planning over the coming years. The general strategy of this approach is to pick an application that is beyond today's state of the art, but that requires well-defined, specifiable advances. Other computer vision applications could be defined that would serve similar purposes; examples are extraction of cartographic features from aerial imagery, robot vision for outdoor construction tasks, and reading of unconstrained cursive handwriting.

Computer vision systems are made up of many components; they make use of image processing algorithms, data structures, and knowledge representations of many different types. Creating a computer vision system to perform a complex task such as vehicle navigation or extraction of cartographic features requires an extensive programming effort involving a team of researchers over a period of months or even years. It also requires major efforts in system integration, an area in which there is little past experience. The level of effort increases still further if the systems must be implemented on a new computer architecture for which a software environment for program development does not yet exist. Building complete vision systems of such advanced types thus can be done only by large research groups and requires substantial funding.

System building efforts can also be undertaken by cooperation among smaller research groups, but there are practical difficulties with this approach. It is often difficult to get researchers to agree on standards for software compatibility, for communication between processes, or even for data formats. Some attention should be given to developing mechanisms that would facilitate and encourage joint cooperative efforts. Competitions between teams of researchers taking different approaches to the same problems would also provide an incentive to produce high-quality results. The competition concept can be applied not only at the level of complete vision systems, but also in connection with smaller research efforts involving specific components of a vision system.

The most effective way to create large experimental vision systems is at a large, well-established research center. Most of the existing centers are associated with universities, but there are several government and corporate laboratories that are also well qualified in this area. It would be desirable to create additional centers of excellence in computer vision, so that experimental system building efforts in a number of different application areas could be pursued. Such centers could also serve as resources for smaller groups. This is particularly important for access to experimental computer systems; ordinarily only a few copies of such systems will exist, and it would be desirable to allow many groups to have access to them.

Computer vision researchers are much sought after by recruiters who want to use their expertise in state-of-the-art applications. It is important to provide incentives that will allow many of these researchers to remain in an academic environment, where they can be free to pursue long-term research goals and where they can also help in the training of the next generation of researchers.

The United States is still at the forefront of computer vision research, but other countries are not far behind. Japan has funded a series of major projects that have had significant impact in the areas of computer vision and pattern recognition. It is important that the United States take similar steps to maintain its lead in the vision field, which has potential payoffs in so many different areas of advanced automation.