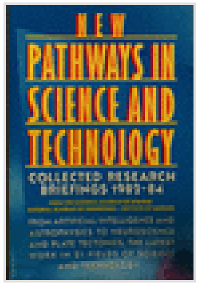

New Pathways in Science and Technology: Collected Research Briefings, 1982-1984 (1985)

Pages
447

Size
5 x 9

ISBN
0309291690

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

NEW PATHWAYS IN SCIENCE AND TECHNOLOGY

NEW PATHWAYS IN SCIENCE AND TECHNOLOGY

COLLECTED RESEARCH BRIEFINGS 1982–1984

COMMITTEE ON SCIENCE,
ENGINEERING, AND PUBLIC POLICY
NATIONAL ACADEMY OF SCIENCES
NATIONAL ACADEMY OF ENGINEERING
INSTITUTE OF MEDICINE

VINTAGE BOOKS NEW YORK
A NATIONAL ACADEMY PRESS/RANDOM HOUSE PUBLICATION

A VINTAGE ORIGINAL, May 1985
FIRST EDITION

NOTICE: The National Academy of Sciences was established in 1863 by Act of Congress as a private, nonprofit, self-governing membership corporation for the futherance 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 PLN-8218330.

Library of Congress Cataloging in Publication Data
Main entry under title:
New pathways in science and technology.
"A National Academy Press/Random House publication."
"A Vintage original"—CIP t.p. verso.
1. Research—United States.
2. Science—United States.
3. Technology—United States.
4. Biology—United States.
I. Committee on Science, Engineering, and Public Policy (U.S.)
II. National Academy of Sciences (U.S.)
III. National Academy of Engineering.
Q180.U5N395 1985 500 84-29153
ISBN 0-394-74061-0

Manufactured in the United States of America

FOREWORD

BY G.A. KEYWORTH, II
SCIENCE ADVISOR TO THE PRESIDENT, AND DIRECTOR, OFFICE OF SCIENCE AND TECHNOLOGY POLICY
EXECUTIVE OFFICE OF THE PRESIDENT

One of the critical responsibilities of anyone charged with making and implementing public policy is to base action on sound principles and informed judgment. As Science Advisor to the President, I have to bring the scientist's perspective to major issues of public policy and I have to make sure that the nation's science and technology resources are developed and maintained to serve a variety of urgent national needs. To do that I must depend on the expertise of the science and technology community for two things: to help anticipate important new avenues in research in order that we can plan for and take advantage of emerging opportunities as soon as possible; and to help establish priorities within technical disciplines in order than we can use limited federal funds wisely.

There are many channels through which expertise is made available to the White House, but none is as comprehensive as that of the National Academy complex*. For the past three years the briefings by the Committee on Science, Engineering, and Public Policy (COSEPUP) have been a valuable resource in meeting our needs for

*The National Academies of Sciences and Engineering, and the Institute of Medicine, are referred to collectively as the National Academy complex.

information. Because the Academy complex has such extraordinary intellectual resources to draw on, the reports in this volume are truly distillations of a tremendous amount of thought and experience within the community.

The subjects of the briefings, which to the reader may seem to have been selected arbitrarily, were strongly influenced by the issues that we in the White House Science Office were working on at the time. While there were obviously many other topics of comparable scientific interest that might have been selected (and many were proposed and debated), the Office of Science and Technology Policy (OSTP) and COSEPUP agreed that the most useful service that the briefings could serve would be to address topical issues in which federal actions were likely in the near term.

The impact of COSEPUP has been in reinforcing perceptions, in strengthening resolve, and in clarifying the often confusing multiplicity of information converging on us. However, I should point out that one of the studies—"Computers in Design and Manufacturing," done in 1983—led almost directly and quickly to an important new program of Engineering Research Centers in the National Science Foundation. The circumstances of timing—COSEPUP's and ours—as well as the willingness of the National Academy of Engineering to respond quickly in developing guidelines for the new program, resulted in a new initiative in a remarkably short time.

Over three years COSEPUP and OSTP have worked well together, and it's my perception that the briefings have become increasingly helpful in formulating science policy. The fascinating lesson of the 1984 round of briefings was the conclusion that serious progress in many of the areas of science and technology of concern to us demand new kinds of multidisciplinary approaches, a conclusion that has stimulated a great deal of thinking about new organizational structures for doing research. Indeed, we may find that this question of structure becomes a dominant theme in coming years as we have to find ways to explore unfolding lines of research in a broader context than the strictly disciplinary methods allow.

Finally, I must acknowledge the contributions of the hundreds of scientists and engineers who worked on these projects, and I want to pay tribute to the man who was largely responsible for the briefings' evolving utility. The late George Low of RPI brought both his wisdom and his leadership to this project for the first two years. He recognized that the purpose of these briefings could only be fulfilled if they addressed current policy issues. His oversight ensured that the briefings would be relevant, concise, and, hence useful. In the past year the COSEPUP leadership has been taken over most ably by Caltech's Lee Silver, and we anticipate a continuing and productive partnership between OSTP and COSEPUP.

CONTENTS

THE REPORTS OF THE RESEARCH BRIEFING PANELS—1984

INTRODUCTION

Contained within this volume are reports from 21 research frontiers of American science and technology. These research briefings were prepared under the guidance of the Committee on Science, Engineering, and Public Policy, of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. Each briefing was presented to senior government offices charged with setting and administering the federal role in science and technology, including the President's Office of Science and Technology Policy, the National Science Foundation, and other agencies.

The briefing reports in this volume, prepared by panels of outstanding researchers in each field, by no means offer a comprehensive view of American science and technology; nor do they constitute an assertion of which research fields are primary. Rather, they reflect issues of immediate concern to the Office of Science and Technology Policy in the year they were done.

With these limits and goals understood, these briefings taken en bloc do serve a wider purpose, one that prompted their publication in this volume. Briefly, they serve both as a montage of American research in the early years of this decade and as testimony to the strength and diversity of U.S. science and technology. Whatever the specific—

THE REPORTS OF THE RESEARCH BRIEFING PANELS

1982

RESEARCH BRIEFING PANEL ON MATHEMATICS

William Browder ***(Chairman),*** *Princeton University*

Herman Chernoff, *Massachusetts Institute of Technology*

Hirsh Cohen, *IBM Watson Research Center*

Louis N. Howard, *Florida State University*

Arthur M. Jaffe, *Harvard University*

Peter D. Lax, *New York University*

Joel Lebowitz, *Rutgers University*

G. D. Mostow, *Yale University*

Michael Rabin, *Harvard University*

I. M. Singer, *University of California, Berkeley*

Shing-Tung Yau, *Institute for Advanced Study*

Guests

Donald Austin, *U.S. Department of Energy*

Stuart Brodsky, *U.S. Department of Defense*

Jagdish Chandra, *U.S. Army Research Office*

David Fox, *Mathematical and Information Sciences, Bolling Air Force Base*

Kenneth Hoffman, *Massachusetts Institute of Technology*

E. F. Infante, *National Science Foundation*

Staff

J. K. Goldhaber, ***Executive Secretary,*** *Office of Mathematical Sciences*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON MATHEMATICS

PREFACE

In response to an invitation from OSTP to present a research briefing on the current state of mathematical research in the United States, a panel was convened on September 25 and 26, 1982. In addition, opinions were solicited from the chairmen of the top 27 research departments of mathematics in the United States.

This document is an account of the deliberations of the panel. Our charge was to identify special opportunities in the mathematical sciences. We could identify many promising areas ripe for development, but it is our belief that the most dramatic mathematical tendency in recent years is the drawing together of mathematics (often the most abstract science) with other sciences and the interplay between them. Some results have already been achieved; with encouragement, the interplay can be made increasingly fruitful.

A healthy mathematical enterprise must have effective means for nurturing new developments. These means involve (1) support of gifted young investigators who will choose from among the new directions and (2) flexibility for scientific leaders allowing them to develop and expose recent breakthroughs. Unfortunately, severe underfunding has limited the mathematical community's capacity to respond in these ways. For this reason, in addition to spelling out the current state of mathematical research, we have tried to specify the additional resources that will be needed to exploit new opportunities.

INTRODUCTION

There is a striking contrast between the importance of mathematical sciences in the United States and the perception of them. On the one hand, mathematics and its applications play an ever increasing role in science, technology, business, and everyday life in this country, and, on the other, mathematical research is almost completely unknown to and poorly understood by the general public and even the scientific public.

The reputation and achievements of the American mathematical community make the United States first in the world in mathematics. Yet at the same time support for mathematical research erodes at a steady rate, and the institutional infrastructure that supports the enterprise exhibits symptoms of decay. In a year in which two out of three of the quadrennial Field's medalists (the mathematical equivalent of the Nobel Prize) are American mathematicians, we find highly ranked departments of mathematics lacking enough research support to send their most productive people to professional meetings or to photocopy important documents. At a time when the development of research in mathematics is making unparalleled progress, when the influence of mathematics is pervasive in the other sciences, when American mathematicians lead the world in most areas, research support for graduate students and young Ph.D.s, the lifeblood of the enterprise, is insufficient to ensure the quality of future generations.

In Section 1, we describe some of the exciting contributions and important developments in mathematical research and the highly promising opportunities that arise from the recent opening of new bridges between mathematics and other sciences. In Section 2, we document the decay of the infrastructure and support of mathematical research, and indicate possible actions to prevent further decay of the mechanisms that have enabled U.S. mathematics to flourish. In Section 3, we recommend steps to be taken to rehabilitate the ailing infrastructure and to provide the flexibility needed for the exploitation of new opportunities.

SECTION 1 SOME RECENT DEVELOPMENTS IN MATHEMATICAL RESEARCH

Our exposition will be in two parts, the first giving examples of the pervasive influence of mathematics in other sciences and the second recounting some of the recent significant advances in theoretical mathematics. Some speculations on future possibilities will be interspersed. (For a somewhat fuller review, we refer the reader to the chapter "On Some Recent Developments in Mathematics" in *Outlook for Science and Technology, The Next Five Years,* NAS, W. Freeman & Co., San Francisco, 1982, pages 467–510.)

INFLUENCE AND APPLICATIONS OF MATHEMATICS

The research area of mathematics most used today in technology is numerical analysis and mathematical modeling. In the area of industrial design, for example, a given process must be described and understood in a mathematical way, and the details of the mathematical description will interact with the design process. Analysis and design become mathematically interdependent. For example, the design of the fuel efficient, weak shock transonic airfoil, currently flying on the Boeing 767, would not have been possible without the mathematical work of Garabedian, Cole, and Jameson. On the speculative side, a project to mathematically model the human circulatory system now under way might eventually have important medical consequences. Among them would be the possibility of indirect ways to measure the heart in a situation where the direct measurement of the heart itself is impractical. Computer-aided design (CAD) is now used in the design of artificial heart valves, using a mathematical model of the left side of the heart.

Mathematical design of efficient compression and turbine blades is a reality today, while the design of efficient combustion chambers is a subject of intense research.

In national defense, the replacement of experimentation by numerical modeling, made possible by advances in computers and dramatic improvements in mathematical algorithms, has resulted in great savings in the cost and improvement in the quality of design. This has been particularly significant in weapons-related research and development, where experimentation is costly, dangerous, and physically impossible in the early stages of a project.

In economics, mathematics is playing an ever increasing role, as witnessed by three recent Nobel prizes in mathematical economics.

In oil prospecting, mathematical results are used in a fundamental way in the separation of primary signals from multiple reflections. The modern theory of inverse scattering is becoming a basic tool in this area. Mathematical modeling is important in the study of efficient secondary oil recovery.

In electrical engineering, the mathematical work of Weiner has proved fundamental in several areas, and mathematical control theory plays an important role.

In medicine, great advances in diagnostic techniques (tomography—the CAT scanner; and NMR) are strongly related to mathematical research. In the latter, methods from singular integrals, complex function theory, and Hilbert space were used. Statistics and statistical methods are crucial in epidemiology, drug testing, and many other areas, and mathematical modeling is an important tool in the development of new drugs. The list could be extended indefinitely with examples drawn from biology, chemistry, neuroscience, and other sciences.

There are many recent examples where mathematical research, driven by the inner dynamic of the subject without reference to practical problems, has been found to be of great significance in other areas. An outstanding illustration of this has been in the development of Gauge Field Theory in physics. Nobel Prize winner C. N. Yang wrote, "I found it amazing that gauge fields are exactly connections on fibre bundles, which the mathematicians developed *without reference to the physical world.*" Algebraic geometry produced all self-dual solutions for the Yang-Mills equations. But the physical theory led also to important consequences in topology, as we will relate later.

Other new and important mathematical inputs into physics have been the introduction of abstract probability into statistical mechanics and material science with the notion of Gibbs state and the input of the theory of dynamical systems and ergodic theory into the study of turbulence. All these phenomena illustrate the drawing together of abstract and applied mathematics and their fruitful interaction.

As a result of the rise of the computer, the theory of computation has become an area of mathematical research. Solidly based on methods and fields in the mainstream of modern mathematics such as probability, combinatorics, algebraic geometry, and number theory, it creates important tools for the practicing computer scientist. The main themes in this new area are the study of algorithms and of programming.

Efficient algorithms often have important practical importance. Notable examples are the Fast Fourier Transform with its application to signal processing and the recently developed randomized algorithms in number theory and finite fields with their application to error-correcting codes and cryptography.

Developments in coding and cryptography provide dramatic examples of unexpected applications of "pure" mathematics to applied areas. Number theoretical work of A. Weil in 1948 was applied to coding theory some years ago. Last year a group of Soviet mathematicians showed how to use the latest work of Deligne, Rapoport, Ihara, and Langlands, in the most abstract areas of algebraic geometry, in the design of error-correcting codes of a theoretical efficiency heretofore deemed impossible.

In the field of robotics, the development of automatic industrial processes depends on successful mathematization or modeling of the processes involved. In many industrial areas, progress is in its infancy, and some of the most simple tasks seem the least likely to yield to automation. It is extremely difficult to design a robot arm with sensors that will enable it to avoid obstructions while picking up a target object, one of the most routine of human abilities. The parameters of this problem can be interpreted as a problem in algebraic geometry, and progress here may have some effect on the solution of other practical problems.

There are current proposals for initiatives in large-scale scientific computing that would have sizable components of mathematical research and important applications to applied mathematics. In pure mathematics, Thurston (one of this year's Field's Medalists) has made a surprising use of the computer as an experimental tool in his work on topology of 3-dimensions, although the solution of the famous four-color map problem a few years ago required the computer essentially in the proof.

Recent advances in computer technology and software are having a deep influence on the nature of work subjected to statistical analysis, on the methods of analysis, and on theoretical questions in statistics. The computer and space technologies provide vast amounts of high dimensional multivariate data which standard classical methods no longer fit, because the underlying assumptions of normality and linearity are no longer satisfied. Methods justified under those assumptions would lead to serious errors. Novel methods of pattern recognition and robust regression and new methods of graphical representation that allow the comprehension of these data are being developed. The interactive kinematic displays of Friedman and Tukey are examples.

New interactive statistical packages with properly built-in diagnostics will permit naive users to observe phenomena formerly accessible only to trained and ingenious statisticians.

Statistics has aided computer science in that simulations and experimental designs have been used to find good hardware configurations and software designs.

PROGRESS IN THEORETICAL MATHEMATICS

By "theoretical mathematics" we mean research motivated by the inner dynamic of the subject rather than by the needs of research in other sciences. It is remarkable how often much of this research, seemingly irrelevant, turns out to have important practical impact. Who, for example, in the 1920s and 1930s would have predicted that the most abstract work in mathematical logic—recursive functions and "Turing machines"—would provide the philosophical framework for von Neumann's introduction of the stored program computer, in which the instructions to the machine can be manipulated and modified by the machine itself? The development led eventually to a multibillion-dollar industry.

Progress has been spectacular in recent years and we cite some notable examples:

The work of Deligne, proving the famous "Weil conjectures" in number theory.

The classification of finite simple groups, as the end of a 20-year effort.

The work of Yau on the Calabi conjecture, with important applications to algebraic geometry.

The work of Thurston, showing how to employ methods of (mostly non-Euclidean) geometry to attack problems in the topology of 3-dimensions.

The work of Khachian on polynomial algorithms in linear programming.

The discovery of solitons and strange attractors.

The work of Connes on operator algebras.

Instead of continuing this list at length, we give in greater detail short accounts of some recent dramatic examples, starting with a startling advance in pure mathematics resulting from interaction with physics.

■ Physicists have introduced gauge theories in 4-dimensions (space-time) as a unifying principle in field theory. The study of Yang-Mills equations of motion in this context led S. Donaldson to a remarkable description of certain 4-dimensional spaces. A little earlier, M. Freedman, using purely topological methods, had produced a powerful comprehensive theory of 4-dimensional manifolds. These results of Donaldson and Freedman have combined to give the following result in the topology of 4-dimensional space. In all other dimensions there is essentially one mode of doing calculus in a Euclidean space (R^n has a unique differential structure for $n \neq 4$), but an entirely different situation exists in dimension 4 (there are at least two different structures on R^4). This qualitative difference between dimension 4 and other dimensions is a startling development for topology, and it may also be the reflection of some deeply significant physical principles.

■ The unifying role of group symmetry in geometry, so penetratingly expounded by Felix Klein in his 1872 Erlanger Program, has led to a century of progress. A worthy successor to the Erlanger Program seems to be Langlands's program to use infinite dimensional representations of Lie groups to illuminate number theory.

That the possible number fields of degree n are restricted in nature by the irreducible infinite dimensional representations of GL(n) was the visionary conjecture of R. P. Langlands. His far-reaching conjectures present tantalizing problems whose solution will lead us to a better understanding of representation theory, number theory, and algebraic geometry. Impressive progress has already been made, but very much more lies ahead.

Closely related to the Langlands program is the remarkable and mysterious connection between counting points in finite spaces and computing the topological invariants of continuous spaces. First propounded in the Weil conjecture, the connection is being made more accessible by the Goreski-MacPherson-Deligne homology theory. The whole thrust of these developments is to force the next generation of mathematicians to embrace heretofore widely separated areas of mathematics. The expected unifications are awesome.

■ In analysis, the old problem of the regularity properties of the Cauchy integral (for Lipschitz curves) was recently solved by the work of Calderón, Coifman, McIntosh, and Meyer. Crucial to the solution of this problem were the techniques of Hardy spaces developed within the last decade, as well as recent methods for dealing with singular integrals with "rough" coefficients. It seems very likely that these ideas will be applicable to a host of important problems in partial differential equations, as is indicated by their role in recent advances in the solution of "Kato's conjecture" (dealing with square roots of Laplacians) and solutions of parabolic equations with minimal smoothness assumptions.

SECTION 2 THE MATHEMATICAL RESEARCH ENTERPRISE

In analyzing the state of mathematical research and its needs, we must keep in mind special features that distinguish mathematics from the other sciences. Among those features are these:

1. Mathematics is the most labor intensive of all sciences. Little equipment is involved, except for computers, which are heavily used in statistics and areas of applied mathematics and are an experimental tool for a few pure mathematicians.
2. The vast majority of research mathematicians are employed in universities as teachers. Industry and national laboratories support only a handful.
3. Very few federal agencies support research in mathematics. NSF supports 60 percent of all research in mathematics and almost 100 percent of pure mathematics; most of the rest is supported by DOD and DOE. This contrasts strongly with other disciplines.
4. The magnitude of the total research support of mathematics by the federal government in comparison with other fields is miniscule, less than $60 million annually.
5. Mathematics is "small science." Though collaboration among 2 or 3 researchers is not uncommon, large projects with many researchers devoted to specific goals are relatively rare. Mathematics thrives on the interaction of independent viewpoints and different approaches.
6. The health of the mathematical enterprise in the United States hinges on the strength and vitality of the departments in the leading research universities.

The following special factors have strongly influenced the pattern of decay in the support of mathematical research that we perceive.

A. It is now generally accepted that the impact of inflation is much greater in labor-intensive enterprises than in the general economy. The impact of declining resources and inflation has, therefore, been most severe in mathematics.

B. The universities as a whole are subject to the same effect, so that the resources available to them as the main supporters of mathematical research have dwindled proportionately.

C. There has been no organization of mathematicians expressing their discipline's support requirements for research. In sciences needing large instruments or projects to achieve their scientific goals, organized support has evolved and served effectively. But in the small-scale individualistic atmosphere of mathematics, no mechanism has evolved for calling attention to the alarming decline in funding.

D. Inflexibility is inevitable when few funding agencies support mathematical research; investigators not supported by NSF, for example, often have no place else to turn, unless their research has clear potential relation to the goals of a mission-oriented agency.

E. The number of top-ranking graduate students seems to be declining, and many of them are from abroad.

F. The strength of some of the leading departments of mathematics is being undermined by the lack of federal funding for research, a lack that the universities cannot replace, especially in states whose economies are suffering.

Yet the small scale of the mathematical enterprise would make it rather inexpensive to alleviate many of the serious shortcomings in the support configuration and to ensure the health and vitality of American mathematical research into the next century.

As the support of mathematical research by federal agencies has eroded over the last decade, the universities have, to some extent, taken up the support, as for example with postdoctoral research instructorships. The support has, however, become more and more difficult for the financially troubled universities to continue.

SUPPORT IN THE MATHEMATICAL SCIENCES

Currently, the United States ranks first in the quality of research in the mathematical sciences. But the vitality conceals a variety of problems that, if left unsolved, will inevitably lead to a substantial deterioration in the nation's mathematical sciences research enterprise. The same can be said substantially of other sciences, but it is our purpose here to document the special strains in mathematics that are reaching the crisis stage.

The consequences of the financial stringency are falling most heavily on the young mathematicians, graduate students, and recent recipients of the Ph.D. Little research support is available to graduate students, though many teaching assistantships are available, particularly at the large state schools. This means that mathematics graduate students, unlike those in other sciences, seldom have the opportunity to work full-time in research. The lack of postdoctoral research appointments in mathematics creates a similar problem for the young Ph.D.s, who, in addition, are finding it increasingly difficult to obtain research grants. *At a time of real opportunities, in the drawing together and mutually fruitful interaction of mathematics and applications, the young innovators who will be needed to exploit the opportunities are not receiving the nourishment they need for full development.*

The chairman of a prestigious mathematics department writes (in a letter to the Panel): "Mathematical research has been flourishing in the past decade, but the institutional structure of mathematical research is in trouble. Recruitment of young talent for the future looks to be in even more serious trouble. The level of research support has been very low in terms of the percentage of active research people supported, and recent cuts in support have produced signs of a serious deterioration of morale especially among younger mathematicians."

Another chairman writes: "We are some one hundred in number, we are invariably ranked among the top twelve departments in the country, we continue to recruit good graduate students, and I claim with confidence that of the one hundred at least ninety are seriously engaged in research and scholarship. Yet, after two severe years, we are down from one-half to about one-third of the faculty on NSF grants. Moreover, we have sustained these severe losses without any sense of the prevalent quality of work having declined at all; on the contrary, several colleagues have lost grants in the very year when they have done their best work. . . . At the same time, universities are increasingly affected by lack of money. Here, for example, loss of NSF grants has reduced departmental income from overhead just when the university, which in any case had always counted on strong departments like ours to earn much of its research support outside, is quite unable to raise the level of state support. Also, of course, we find little endowment money coming in earmarked for mathematics."

Yet another writes: "Many young mathematicians are discouraged at their prospects for a successful career in mathematics because of decreased research funds, poor salaries and the shortage of openings in universities. Several I know are actively looking for jobs in other fields where they can expect much better treatment economically, and even the top departments are finding it increasingly difficult to attract qualified graduate students."

What is the research support picture in the mathematical sciences? One can glean an indication from Tables 1 and 2.

Lest the disparity in funding exhibited in Table 2 be totally attributed to differences in the numbers of professionals in the various areas, we note that in January 1980 the numbers of full-time scientists and engineers at doctorate-granting institutions in the various areas were as follows:

Engineering—20,511
Life Sciences—93,309
Physical Sciences—16,845
Environmental Sciences—5,891
Mathematical Sciences—9,146*

In light of these data, it is not surprising that departmental chairmen speak of discouragement and of deterioration of the morale of young mathematicians. If it is in the national interest to maintain a healthy and vigorous mathematical sciences research enterprise, then it is imperative that conditions contributing to this deterioration be altered. We discuss the more important problem areas, and for each area estimate the dollar cost of alleviating the problem.

POSTDOCTORAL POSITIONS IN THE MATHEMATICAL SCIENCES

Table 3 (based upon data in "Academic Science: Graduate Enrollment and Support for 1980," NSF 81-330, Table A-30) gives dramatic evidence of the disparity in the numbers of postdoctorates in various sciences and in engineering.

The excellence of science in the United States today derives from postdoctoral opportunities in the past. Clearly, if the current postdoctoral pattern in the mathematical sciences persists, we will jeopardize the quality of the mathematical sciences at our leading universities in the years to come. To be sure, the pool of outstanding candidates for tenure positions at the leading three or four universities will be large

*Data from "Academic Sciences: Scientists and Engineers," NSF 81-307, Table B-5. These figures include both research and non-research scientists and engineers.

TABLE 1

	Percentage of Research in Universities Not Sponsored November 1978–October 1979
Engineering	16
Environmental Sciences	16
Life Sciences	13
Mathematical/Computer Sciences	59
Physical Sciences	22

Source: "Activities of Science and Engineering Faculty in Universities and 4-Year Colleges 1978–1979; Surveys of Science Resources Series NSF Final Report," NSF 81-323, page 2.

TABLE 2

	Estimated Federal Obligations for Research* Performed at Universities and Colleges, Fiscal Year 1981
Engineering	$ 350,208
Environmental Sciences	$ 332,063
Life Sciences	$2,088,893
Mathematics	$ 55,906
Physical Sciences	$ 511,638

*"Research" does not include "R&D Plant," which is defined as follows: "R&D plant (R&D facilities and fixed equipment, such as reactors, wind tunnels, and radio telescopes) includes acquisition of, construction of, major repairs to, or alterations in structures, works, equipment, facilities, or land, for use in R&D activities at federal or non-federal installations. Excluded from the R&D plant category are expendable equipment and office furniture and equipment. Obligations for foreign R&D plant are limited to federal funds used in support of foreign research and development." Ibid, page 2.

Source: "Federal Funds for Research and Development, Fiscal Years 1979, 1980, 1981," NSF 80-318, page 117.

enough. However, one can expect a serious drop in the quality of candidates at the next five ranking universities, and an even more serious drop at the next ten and twenty.

There is a clear need to provide a significant number of outstanding recent Ph.D.s in the mathematical sciences with the opportunity

TABLE 3

	Number of Postdoctorates in All Graduate Institutions: 1980
Engineering	981
Environmental Sciences	311
Life Sciences	11,715
Mathematics	143
Physical Sciences	4,261

to devote full time to research in association with a major scientific figure of their own choosing. The most creative future mathematicians in the United States will emerge from this group, and it must be nurtured. We are not advocating a large move away from teaching, the traditional mode of mathematical support. Typically, all graduate students will do some teaching while preparing for their doctorates. However, we are suggesting a small shift to allow promising young investigators a few years after their degrees to develop their research talents.

The Panel estimates that there should be an additional 120 postdoctoral appointments each year, each appointment being for two years. At a cost of $25,000 per appointment per year, this amounts to an increment of $6 million per year. It is also essential, in the view of the Panel, that there be flexibility in the nature of the postdoctoral support. A variety of modes should be used in offering it: fellowships in NSF, DOE, DOD; institutional support at major centers and research institutes; enhancement of grants by providing support for postdoctoral positions.

RESEARCH GRANTS IN THE MATHEMATICAL SCIENCES

Table 4 indicates many categories where funding for the mathematical sciences is markedly insufficient. The amount of dollar support available for graduate students in the mathematical sciences is inadequate. We illustrate this with a very particular example. A member of this Panel recently directed two Ph.D. dissertation students, one in the mathematical sciences, the other in another science. The non-mathematics student was federally supported, and his sole task was to devote

TABLE 4
BUDGETARY CATEGORIES OF NSF AWARDS: FY 1981

BUDGETARY CONSTRAINTS	Mathematical Sciences AMOUNT	Mathematical Sciences PERCENTAGE OF TOTAL	MPS Directorate[a] AMOUNT	MPS Directorate[a] PERCENTAGE OF TOTAL
Personnel:				
Senior Personnel	$11,710	41.6	$ 23,253	10.2
Postdoctoral Associates	1,699	6.0	17,068	7.5
Graduate Students	1,189	4.2	26,611	11.7
Other Personnel Costs[b]	2,538	9.0	29,027	12.7
SUBTOTAL: Wages, Salaries, and Fringe Benefits	17,136	60.8	95,959	42.1
Permanent Equipment	50	.1	38,550	16.9
Other Direct Costs[c]	2,022	7.2	41,141	18.0
Indirect Costs	8,972	31.9	52,370	23.0
TOTAL: All Budgetary Categories	$28,180	100.0	$228,020	100.0

[a]Excludes mathematical sciences section, MCS.
[b]Includes undergraduate students, secretarial-clerical, other professionals, technicians, and fringe benefits.
[c]Includes domestic and foreign travel, materials and supplies, publication costs, consultant costs, and computer costs.

his time to study and research. The mathematics student, on the other hand, was university supported and, in addition to study and research, had to devote his time to grading papers, teaching, registering students, and holding office hours. The difference in treatment did not go unnoticed by these students nor, assuredly, did others fail to notice it. Mathematics students need the opportunity for a year or two of uninterrupted research during their graduate study to fully develop their research abilities, as students in most other fields of science do.

Along with decreased research support there is a rapid increase in the demand for mathematics courses in universities. Those who enter and remain in the mathematical sciences must necessarily teach more and devote less time to research. This is now characteristic of the mathematical sciences but not of most of the other sciences.

Because of this dearth of support and less than optimal conditions for the acquisition of knowledge in mathematics, many talented students select other areas for study. As a consequence, the quality and excellence of the graduate student body in the mathematical sciences are diminishing, and, as with the postdoctorates, the implications for the future excellence of the discipline in the United States are ominous.

Implicit in Tables 1 and 4 is that a substantial amount of support for research in the mathematical sciences must be contributed by universities in the form of reduced teaching loads, secretarial services, travel costs, and other aids. This support, as we have noted, is crumbling.

It is the view of the Panel that an additional 300 graduate students should be supported per year. At a cost of $20,000 per student (including indirect costs) this amounts to an increment of $6 million per year. In addition, the amount allocated in federal mathematical sciences research grants for secretaries, travel, and publication should be increased by $3,000 per individual investigator. This amounts to an increment of $5.4 million per year.

PERCENTAGE OF ACTIVE RESEARCH PEOPLE SUPPORTED

We saw in Table 1 that, in fiscal year 1979, 59 percent of the mathematical sciences research done in universities was unsponsored. Since then, the situation has deteriorated even further: *approximately 200 active researchers doing high-quality work have lost support during the past two years. Moreover, the research of many excellent new Ph.D.s in the mathematical sciences—more than 86 percent—goes unsupported.* We are not capitalizing on the investment made in the development of mathematical scientists.

To stem the decay and to put a measure of vitality into the mathematical sciences research enterprise require the allocation of sufficient funds to support an additional 500 researchers. At a cost of $20,000 per researcher, including indirect costs, this amounts to $10 million per year.

STATE OF THE INFRASTRUCTURE

The stresses addressed thus far pertain primarily to those that affect the individual researcher or graduate student. Severe problems also exist in the area of communication and interaction between researchers. There is a paucity of mechanisms in the mathematical sciences for maintaining the vitality of researchers working at a distance from elite centers, for generating young people's interest in promising and important new subfields, and for informing and educating the research community about new ideas and results—particularly those ideas and results that lie at the boundary of two, or more, scientific disciplines. Those mechanisms that do exist are in disrepair. Support here is especially important because of the new opportunities made possible by the recent liaisons between mathematics and other sciences.

There has been substantial discussion in the mathematical sciences community concerning means of supporting the infrastructure of the research enterprise. In Table 5 we list some of these means and the amount of funds that the Panel feels will be required per year to support them.

Table 6 is a summary of the per annum dollar amounts discussed above (*not* listed in priority order).

SECTION 3 RECOMMENDATIONS

The mathematical sciences lie at the core of science, technology, and the national defense. An excellent mathematical sciences enterprise does not automatically produce excellence in science and technology and strength in the national defense, but one cannot have quality and strength in the latter without quality and excellence in the mathematical sciences.

The cost to the federal government of ensuring excellence in the mathematical sciences is relatively very small, the leverage of the dollars invested very large, the ratio of benefit to cost enormous.

In times of economic uncertainty or stress, the implementation even of programs with high benefit-to-cost ratios is sometimes delayed. When this occurs, one must also look at the "disbenefit" associated with delay. The current exciting opportunities will not be fully exploited in the United States; the erosion of excellence in the mathematical sciences will accelerate; the disbenefit to our nation of not doing something now is too large to allow the erosion to continue.

The Panel makes the following recommendations:

1. *The federal dollar allocation for research in the mathematical sciences should be increased over the next three years by approximately 80 percent—i.e., there should be a total increment over the next three years of $42.8 million (in 1982 dollars).*

TABLE 5

Summer schools, special years, mini-institutes		$ 5.3 million
Mid-level fellowships		3.6 million
Travel grants, senior research associate programs		4.5 million
Computer time and equipment		2.0 million
	TOTAL	$15.4 million

TABLE 6

Postdoctoral positions		$ 6.0 million
Graduate students		$ 6.0 million
Operating expenses in grants		$ 5.4 million
Increase in number of grants		$10.0 million
Infrastructure (total)		$15.4 million
	TOTAL	$42.8 million

2. *The increments should be allocated in proportions deemed appropriate to the NSF, the DOD agencies, and the DOE.*
3. *The managers of mathematical science funding programs should have sufficient flexibility and freedom to choose the areas of research to be supported and the mechanisms for support.*

The Panel recognizes that it has no formal status as representative of the mathematical community. No single group does. Nevertheless, because it realizes that increased support may be slow in coming, despite the emergency, the Panel tried to give priority to its recommendations as follows.

■ If there is no increase, we recommend no changes. The mathematical community, after deliberation, has recently reallocated the resources available to it. It shifted monies into an alternative mode of research support and cut down on the number of individual research grants. It will take time to absorb these changes. We must emphasize again, however, that no increase in support will spell disaster for all but a few of the top mathematical research centers.

■ If there is a 10 percent increase ($5.7 million), we recommend that the increment be disbursed approximately as follows:

Allocation	Use
$1.5 million	Postdoctoral Positions (in a variety of modes)
$1.2 million	Increasing the Operating Expense Allocations in Grants
$0.5 million for each of the following:	Mini-Institutes Senior Research Associate and Visiting Positions in Grants Mid-level Fellowships Graduate Student Support Equipment and Computer Time
$0.25 million for each of the following:	Increasing the Number of Grants Travel Grants, Special Years, etc.

■ If there is a 20 percent increase ($11.5 million), we recommend that the increment be disbursed approximately as follows:

Allocation	Use
$2.0 million	Increasing the Operating Expense Allocations in Grants
$2.0 million	Postdoctoral Positions (in a variety of modes)
$1.75 million	Increasing the Number of Grants
$1.0 million	Senior Research Associate and Visiting Positions in Grants
$1.0 million	Graduate Student Support
$0.75 million for each of the following:	Mini-Institutes Mid-level Fellowships Research Institute Equipment and Computer Time Travel Grants, Special Years, etc.

■ If there is a 50 percent increase ($28 million), we recommend that the increment be disbursed approximately as follows:

Allocation	Use
$4.8 million	Increasing the Number of Grants
$4.6 million	Postdoctoral Positions (in a variety of modes)
$4.3 million	Increasing the Operating Expense Allocations in Grants
$3.5 million	Senior Research Associate and Visiting Positions in Grants
$3.2 million	Graduate Student Support
$2.2 million	Mini-Institutes
$1.7 million	Mid-level Fellowships
$1.4 million	Equipment and Computer Time
$1.4 million	Research Institute
$0.9 million	Travel Grants, Special Years, etc.

■ If there is an 80 percent increase ($42.8 million), we recommend that the increment be disbursed approximately as follows:

Allocation	Use
$ 6.0 million	Postdoctoral Positions
$ 6.0 million	Graduate Students
$ 5.4 million	Operating Expenses in Grants
$10.0 million	Increasing the Number of Grants
$ 5.3 million	Summer Schools, Special Years, Mini-Institutes
$ 3.6 million	Mid-level Fellowships
$ 4.5 million	Travel Grants, Senior Research Associate Programs
$ 2.0 million	Computer Time and Equipment

RESEARCH BRIEFING PANEL ON ATMOSPHERIC SCIENCES

Charles L. Hosler, Jr. ***(Chairman),*** *Pennsylvania State University*

Richard A. Anthes, *National Center for Atmospheric Research*

Louis J. Battan, *University of Arizona*

Ralph J. Cicerone, *National Center for Atmospheric Research*

Peter V. Hobbs, *University of Washington*

John E. Kutzbach, *University of Wisconsin, Madison*

C. Gordon Little, *NOAA Wave Propagation Laboratory*

Michael B. McElroy, *Harvard University*

Norman A. Phillips, *National Meteorological Center (NOAA)*

Joseph Smagorinsky, *Princeton University*

Verner E. Suomi, *University of Wisconsin, Madison*

Staff

John S. Perry, ***Executive Secretary,*** *Board on Atmospheric Sciences and Climate*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON ATMOSPHERIC SCIENCES

SUMMARY

Advancing scientific understanding and technology have put us on the threshold of dramatic improvements in our ability to predict the future state of the atmosphere. At this time, special opportunities for scientific advances, which also involve important social concerns, are presented by three areas of the atmospheric sciences: small-scale severe weather systems, atmospheric chemistry, and climate.

This brief report recommends specific actions that should expand our knowledge and predictive capability in each of these areas. The recommendations are based upon the scientific merits of the various actions and do not consider details of execution and cost. Advances in these three areas will improve human productivity and efficiency, protect life and property, and enhance the quality of life. (See Table 1.)

INTRODUCTION

The global atmosphere forms an indivisible, unified object of study: space-borne observation/communications platforms to view the earth as a whole and cooperative programs among all the world's nations are

TABLE 1
HIGH-LEVERAGE INVESTMENTS IN THE ATMOSPHERIC SCIENCES

Problem Area	Recommended Action	Probable Benefits
1. Small-scale Severe Weather Systems	Use near-term technological enhancements (Doppler radar and wind profiling, data management and display) as a basis for planning long-term mesoscale research	Immediate improvements in forecasts/warnings; improved aviation safety; basic understanding; accelerated progress in research; improved prediction techniques
2. Atmospheric Chemistry:		
a. Acid Rain	Conduct field studies on dry deposition, in-cloud processes, and remote-region sources	Cost-effective and environmentally sound control strategies
b. Tropospheric Chemistry	Support a comprehensive research program, including data base	Air quality protection and climate prediction
c. Stratospheric Ozone	Support the Upper-Atmosphere Research Satellite (and associated research)	Ozone layer protection
3. Climate Variability and Change		
a. Long-term Climate Change	Study the circulation of the World Ocean and its interaction with the atmosphere and climate Measure and model trace gas concentrations	Predictive models for long-term climate change Assessment of effects on radiation balance and climate
b. Short-term Climate Variations	Support El Nino/Southern Oscillation studies Conduct prediction studies with general circulation models	Seasonal/interannual predictions Monthly/seasonal predictions

fundamental and indispensable prerequisites for meaningful research. Thus, while the atmospheric sciences, like other branches of the scientific enterprise, continually present challenging problems and opportunities, their progress depends to an unusual extent upon multidisciplinary investigations, a global supporting infrastructure of space- and ground-based observing/monitoring programs (largely justified by operational requirements), and multinational activities of many kinds.

Man's influences on certain atmospheric processes are now comparable in some respects to nature's. By release of carbon dioxide, industrial effluents, and entirely new substances such as the chlorofluorocarbons, we may be changing the radiative and chemical nature of the atmosphere and thus affecting its ability to maintain a stable climate and a habitable world for humans.

Through daily weather, through the slow march of climate, through effects and control of pollution, the atmosphere influences myriad daily activities and decisions throughout our complex society, all of which contribute to our nation's efficiency and productivity. The atmosphere's constituency thus includes not only pilots, farmers, commodity dealers, and builders, but every citizen. To ensure our continued welfare in a variable and vulnerable atmosphere, we must acquire better fundamental understanding of the atmosphere and an enhanced capability to predict its changes arising from both natural processes and human interventions.

SMALL-SCALE SEVERE WEATHER SYSTEMS

During the past decade there has been a growing consensus within the meteorological community that significant improvements can be made in the short-range forecasting (0–12 hours) of mesoscale—or stormscale—weather systems. In terms of size and duration, these lie between individual clouds and the larger storms depicted on the conventional weather map. Although they tend to elude the conventional observing/forecasting system, they produce much of our weather and most of our severe weather: tornadoes, flash floods, hailstorms, snowstorms, and strong winds. Optimism about our ability to forecast these systems is based on recent advances in both technology and understanding. Technological improvements include satellite systems (which provide information on cloud cover and temperature and humidity soundings), various types of radars (which can provide continuous and quantitative information on winds, precipitation, and—

through pattern recognition—various weather phenomena), and electronic means for the rapid assimilation of this large quantity of information and its communication to the public in the form of timely forecasts. Scientific advances have been made in the understanding of such phenomena as frontal circulations and precipitation, severe mountain winds, and severe thunderstorms which produce hail, wind shear, and sometimes tornadoes.

Currently, much of this technology is not available to the public or private weather forecaster, although it has been well tested in the research community. Transfer of the technology to the forecasting sector should enable weather services, public and private, to make more timely and accurate predictions of stormscale weather phenomena. Such improvements would aid in protecting life and property, serve the national economy, and help meet defense requirements. Widespread employment of this technology would also provide a badly needed research data base and accelerate the progress of research in this area.

LONG-TERM MESOSCALE RESEARCH

Two areas should be emphasized in long-term mesoscale research. First, we must improve our understanding of the physical processes that determine the growth, maintenance, and decay of mesoscale weather systems. This improved understanding can be achieved through appropriate special observational programs and theoretical research. Such knowledge is essential for the development and improvement of the conceptual and numerical models necessary to improve the forecasts and warnings of significant mesoscale weather.

The second major research objective is to develop methods of analyzing the massive data sets that would be available from the new measuring facilities. The development of analysis schemes and the communication systems required to process, transmit, and display the data are necessary both for short-range forecasts and warnings and for numerical weather prediction models. The recommendations outlined below will lead the way to more effective research in these areas, and will also begin to realize the improvements in short-range severe weather forecasting made possible by new technology.

NEAR-TERM TECHNOLOGICAL ENHANCEMENTS

Rapid exploitation of several technological advances could quickly accelerate our research progress and improve our forecasting capabilities. Doppler radar operating in the near-horizontal can detect the special motion patterns within thunderstorms that precede the formation of tornadoes. Also, the formation of weather-producing stormscale circulations, such as squall lines or heavy snow bands, can be detected from small changes in wind patterns before they have formed. Tests have demonstrated that Doppler radar observations can lead to earlier issuance of tornado warnings and an appreciable reduction in false alarms, thus enhancing public confidence in and response to warnings. Such radars can also warn of thunderstorm-generated severe downdrafts and wind shears that pose hazards to aircraft in landing and takeoff. Federal recognition of this potential already exists in the interagency NEXRAD program, involving the Department of Defense, the Federal Aviation Administration, and the National Weather Service. Funding for this program should be accelerated in order to keep all essential elements of the program moving on schedule. Full financial support of this system presents the most effective contribution to improved forecasting of severe weather that the federal government can now bring about. The routine use of Doppler radar will also increase scientific understanding of these destructive and erratic wind storms and the small-scale circulations hazardous to aviation.

■ *In the near term, the technological advances offered by Doppler radar should be exploited by accelerating funding of the interagency NEXRAD program.*

Radar wind profiling systems using longer wavelengths can measure upper-air winds to considerable heights from naturally occurring refractive index variations. Radiometric temperature sounding systems can similarly obtain temperature and pressure data. The systems operate unattended and produce accurate and continuous observations in all weather conditions. The latter characteristics are completely new and highly valuable from both scientific and operational viewpoints. Continuous observations permit accurate tracking of significant characteristics of the flow of weather, such as fronts and wind shear lines. The continuous availability of enroute winds for aircraft would also make possible significant fuel savings. In addition, dual-polarization Doppler radars can provide valuable information on cloud and precipitation processes for research purposes.

■ *An array of Doppler radar upper wind sounding systems should be deployed promptly in parts of the country most subject to violent weather to permit evaluation of their impact on weather forecasting and to aid research on mesoscale weather systems.*

Both research and applications on the storm scale are currently limited by the coarse horizontal and temporal resolution of our present upper air data network and data handling systems. Radar can provide additional information, and sounding instruments on geostationary satellites are beginning to provide immense improvements in horizontal and temporal resolution. In addition, frequent images in the visible, the infrared, and water vapor bands are becoming available from satellites. Interpreting this enormous flood of data has been likened to trying to drink from a fire hydrant without drowning. Over the last several years, however, demonstrations have shown that it is possible to digest this data flood in real time and to convert it into highly useful information for forecasting and research. To do this, the user must be provided with computer-based video displays and graphics capabilities. When this is done, the improvement in forecasting performance has been dramatic. Unfortunately, our existing system for communicating and displaying this information to the user falls far short of what is possible. The communications industry continues to develop an impressive number of new communications systems, including direct broadcast from satellites, that could serve to collect and disseminate weather information and support services to a wide range of users including researchers, forecasters, and the general public. This is an unusual opportunity to provide large improvements at reasonable cost.

These capabilities can be and should be made available in the near future for use by the weather services and the research community. Their widespread availability would provide immediate economic advantages and reductions in losses of lives and property. Moreover, they would provide the basic framework needed for a longer-term program of research on small-scale weather systems. In addition, the availability of state-of-the-art data management and display capabilities to the research community would significantly accelerate progress in other important areas such as climate.

■ *Federal agencies should promptly initiate actions to make advanced communications, data management, and display capabilities available both to the research community and to operational forecasters.*

The above near-term technological enhancements would provide the basis for development of a long-term research program to improve our scientific understanding and operational capabilities in small-scale severe weather systems.

ATMOSPHERIC CHEMISTRY

The past decade has seen a revolution in the field of atmospheric chemistry. In the 1970s, chemists were primarily engaged in identifying problems and reacting to crises; in the 1980s they are seeking a new level of maturity for this area of science. They have learned that crisis assessment, in the absence of a deep understanding of the chemistry and physics of the natural atmosphere, is a hazardous exercise.

In a larger context, the earth is a planet characterized by change. Moreover, the human race has now achieved the ability to alter its environment on a global scale, with implications for food production, the quality of air and water, and the integrity of the global chemical cycles essential to life. In the past, human needs could be met by the expansion of frontiers, by land clearance, by the application of chemical fertilizers and pesticides, by irrigation, and through the exploitation of energy resources harvested from the sun by the biosphere millions of years ago. An understanding of the overall system is essential if the human race is to live successfully with global change.

Several current issues relating to atmospheric chemistry are singled out below for special near-term emphasis.

ACID RAIN

Many of the general problems facing atmospheric chemistry are illustrated by the acid rain problem. Unlike Europe, no routine monitoring of the chemical properties and acidity of precipitation has been conducted in North America until very recently, and the effects of increased industrialization on the acidity of precipitation have been largely undocumented. Similarly, little is known about the chemical composition and acidity of "natural" (i.e., "clean" background) rainfall in remote regions of the world, which sometimes exhibit unexpected acidity. Since the cost of controlling industrial emissions is large, and the economic consequences could be great, the effects of control strategies on acid rain should be carefully assessed scientifically.

The principal contributors to acid rain are compounds of sulfate and nitrate, both of which are released by human activities. These compounds may appear in precipitation through two mechanisms: "scavenging" of sulfate and nitrate compounds already present in the atmosphere and/or "*in situ*" production through incorporation of trace gases into cloud and precipitation droplets followed by chemical reactions within the droplets to form sulfates and nitrates. Acidic material also reaches the ground through dry deposition. Establishment of the dominant mechanism ("scavenging," "*in situ*" production, or dry deposition) will dictate, in part, appropriate control strategies; it might even indicate that currently available control techniques would have relatively little effect on the occurrence of acid rain.

■ *Three essential investigations should be undertaken in an attack on the problem of acid rain: Field measurements of dry deposition for acidic gases, e.g., SO_2 and HNO_3; field studies of in-cloud chemistry and cloud physics from aircraft or at mountain laboratories; and assessment of natural sources of acidity in remote regions.*

TROPOSPHERIC CHEMISTRY

An understanding of the acid rain problem will require a broad assessment of tropospheric chemistry, not simply on a local but also on a global scale. For example, a major gap has developed in the measurement and understanding of tropospheric ozone. Available data and theory indicate that commercial aircraft operations can increase tropospheric ozone. Photochemical pathways to ozone production in the troposphere should be simpler than those in urban smog, and measurements of the relevant variables should be easier. Coordinated measurements are needed, techniques need to be developed, and global measurements from space will eventually be required to provide a data base for global tropospheric models.

More generally, tropospheric chemistry has a major impact on the quality of life for the planet as a whole, e.g., air quality and climate stability. The natural state of the tropospheric system, including its interaction with the global scale biosphere, must be defined to provide a basis for assessment of such issues as aid rain. A long-term integrated program of observations and research is now being defined by a study group of the National Research Council.

■ *Acquisition of a data base and a level of understanding sufficient to define the natural state of the troposphere should be a major goal for an integrated program of tropospheric research. To this end a program of research should be mounted to develop data on the distribution of important gases such as ozone, and a strategy should be implemented to improve our understanding of chemical cycles, with definition of important biospheric sources and sinks for key species.*

STRATOSPHERIC OZONE

Scientific interest is now aimed at improving basic understanding of the stratospheric photochemical system. It is possible that the redistribution of stratospheric ozone caused by man's activities will be as important environmentally as the change in the total amount of ozone. Fuller quantitative understanding of man's impact on the ozone layer is needed to provide a sound basis for future decisions by industry and government.

Observations are needed to test available models of stratospheric chemical and dynamic processes and to guide the development of future models. Strategies for an orderly study of the stratosphere have been developed involving all segments of the concerned scientific communities—atmospheric chemists, laboratory chemists, meteorologists, and modelers. The research is interdisciplinary and international in scale, and it demands the collaboration of government, academia, and the industrial sector.

■ *The Upper Atmospheric Research Satellite (UARS) is a central element of this nation's stratospheric research program, and a commitment to this satellite program should be made at the earliest possible time.*

CLIMATE

In the past dozen years, the attention of atmospheric and oceanic scientists has increasingly turned to problems of climate. Quantitative study of past climates has revealed the long-term variability of climate, while the Sahel drought and the grain trade dislocations of the early 1970s showed our vulnerability to changing climate. Numerical models and field observations (e.g., the measurement of increasing atmospheric concentrations of carbon dioxide and some other radiatively active trace gases) raised concern that our own actions might be producing damaging and irreversible changes in climate. However, it was realized that both prediction of natural climate variations and assessment of man-made changes require as an indispensable prerequisite a basic understanding of the workings of the global climate system comprising the oceans, the land surface, the ice and snow masses, and the atmosphere, and the ability to construct quantitative models based upon this understanding. These concerns prompted a worldwide surge of research on the global climate system and the organization of coordinated national and international programs. They also focused interdisciplinary research effort on critical problems such as the increase in atmospheric carbon dioxide. These endeavors are well under way and should proceed unabated. However, they have already revealed several opportunities that merit exploitation.

WORLD OCEAN CIRCULATION

The World Ocean, covering 70 percent of the planet's surface, is an integral component of the climate system and provides its long-term memory. The dynamics of the oceans, however, are only poorly understood in comparison to those of the atmosphere, impeding our ability to model the coupled combined systems. Hence, we are unable to determine adequately the important thermal and chemical interactions between the ocean and the atmosphere. As a result, we cannot predict climate beyond a season or elucidate completely the sensitivity of climate to a CO_2 increase.

A fundamental necessity is to establish observationally the three-dimensional structure of the circulation of the World Ocean, its seasonal variation, its interannual variability, and its interactions with the atmosphere. Only then can we establish definitively the relative role of the ocean in storing and transporting heat in the climate system, stimulate modeling improvements, and supply a validation base for climate simulation studies. Such advances would be of equal benefit to the discipline of oceanography.

New technology for ocean measurements makes it possible to conceive of an observational experiment to determine the circulation of the World Ocean. Satellite-borne radar altimeters can determine the sea-surface topography, while scatterometers coupled with other observations provide measures of wind stress. Acoustic tomography promises ocean-scale density measurements, and measurements of man-made tracers directly reveal long-term circulations. These and other observations will supplement the ongoing global observing system.

■ *Investments should be begun now to ensure the availability of long lead-time satellite-borne observing and communications capabilities (typified by the TOPEX program) to permit initiation of a world ocean circulation experiment by the end of this decade.*

TRACE GASES

Much attention has been focused on projected global warming due to increased atmospheric CO_2. However, the corresponding effects of other radiatively active trace gases are additive to those of CO_2 and can be significant. Concentrations of nitrous oxide, the chlorofluorocarbons, methyl chloroform, and methane have increased worldwide, and there is circumstantial evidence that tropospheric ozone is also increasing. These and other gases absorb infrared energy in otherwise transparent regions of the infrared spectrum and thus block heat from leaving the earth.

Monitoring the concentrations of these gases (except for ozone) is economical and feasible. Vertical profiles with good latitudinal coverage are needed for nitrogen oxides, ozone, some hydrocarbons, water vapor, and carbon monoxide, among others, as a basis for theoretical and modeling improvements.

■ *Field and laboratory measurements should be taken to provide a firmer basis for our understanding and prediction of the behavior of the trace gases that influence climate.*

EL NINO, THE SOUTHERN OSCILLATION, AND INTERANNUAL PREDICTABILITY

A large and growing body of evidence suggests that the tropical oceans influence the global circulation of the atmosphere on time scales from a season to about two years and provide encouraging possibilities for prediction of seasonal and interannual climate fluctuations.

The Southern Oscillation encompasses a large variety of phenomena throughout the global atmosphere and ocean that are known to be correlated with the large-scale oscillations of south Pacific surface pressure to which the term was first applied. These phenomena include large increases in water temperature off the Peruvian coast (El Nino), equatorial sea-surface temperature anomalies, variations in the Indian Monsoon, and seasonal climate irregularities over North America. A full cycle of the Southern Oscillation endures over about a two-year period, and is initiated aperiodically at intervals of two to ten years. It is the largest coherent signal in short-term climate variability.

Much of what has been learned recently about the interannual variability of the tropical ocean and the global atmosphere has come from existing observational sources and climate models, and the implications of these findings for climate predictability should be exploited promptly. Already, simple empirical methods based on these phenomena are being applied to produce crude seasonal forecasts over North America.

Further progress will require a better understanding of atmospheric and, especially, oceanic mechanisms. A more complete description of the chain of events is needed. What are the energetics of the cycle? What triggers it? Is there a beginning to the sequence? What determines the length of the cycle? Why is the frequency of occurrence so irregular?

A strong base for productive research exists in a group of Pacific-oriented oceanographic programs completed, in progress, or already planned. Study of this problem is now under way, and a national plan for a comprehensive research program in both the atmosphere and the ocean will be completed before the end of 1982. International interest and participation are already evident.

■ *A concentrated program of observation and research should be mounted in the early to mid-1980s in strong and continuing support of El Nino/Southern Oscillation studies.*

MONTHLY AND SEASONAL PREDICTION

Recent research with global atmospheric general circulation models designed for prediction in the 5- to 15-day range indicate that in some instances atmospheric events may be predictable in the 30-day range. In particular, blocking episodes* and their attendant dramatic regional temperature and rainfall anomalies and shifts in storm tracks have been successfully predicted. There is some evidence that this enhanced predictability could be extended to the seasonal range. The investigation of atmospheric predictability was a prime objective of the Global Atmospheric Research Program, in particular of the 1979 Global Weather Experiment. This highly successful international endeavor yielded a unique body of detailed data on the time-dependent behavior of the atmosphere. Success in this research will require the sustained availability of substantial research funds.

*The term "blocking" refers to the development of strong and almost stationary circulation patterns that persist over large regions for extended periods, i.e., a few days to several weeks.

■ *Comprehensive observational and theoretical studies should be conducted to discover the physical reasons for persistent circulation states and to determine whether there are any other circulation characteristics that may also be predictable in the monthly/seasonal time range.*

RESEARCH BRIEFING PANEL ON ASTRONOMY AND ASTROPHYSICS

George Field *(Chairman)*, *Harvard-Smithsonian Center for Astrophysics*

Geoffrey Burbidge, *Kitt Peak National Observatory*

George Clark, *Massachusetts Institute of Technology*

Arthur Code, *University of Wisconsin*

Carl Fichtel, *NASA Goddard Space Flight Center*

William Hoffman, *University of Arizona*

Kenneth Kellermann, *National Radio Astronomy Observatory*

Jeremiah Ostriker, *Princeton University*

Tobias Owen, *State University of New York, Stony Brook*

Harlan Smith, *University of Texas, Austin*

Harvey Tanabaum, *Harvard-Smithsonian Center for Astrophysics*

Arthur Walker, *Stanford University*

Staff

Paul Blanchard, ***Rapporteur*** *(Harvard University)*

Bruce Gregory, ***Executive Secretary,*** *Space Science Board*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

5

REPORT OF THE RESEARCH BRIEFING PANEL ON ASTRONOMY AND ASTROPHYSICS

I. ASTRONOMY AND ASTROPHYSICS FOR THE 1980'S

We live in an extraordinary age for astronomy, unmatched since the time of Galileo. The discoveries of the past 20 years, made from both ground-based and space observatories, have radically changed our understanding of the Universe. Quasars, x-ray sources, pulsars, polyatomic molecules in interstellar space, and the cosmic microwave background radiation were discovered in the 1960's. The 1970's added discoveries of accretion of matter by neutron stars, observational evidence for a stellar black hole, intergalactic gas at 100 million degrees, apparent expansion at greater than the speed of light in quasars, volcanic activity on Io, the origin of high-velocity solar-wind streams in "holes" in the solar corona, rings around Uranus and Jupiter, observational evidence for gravitational radiation from a binary star, gamma-ray bursts, and the gravitational refraction of light from quasars by intervening galaxies.

Through federal support, U.S. astronomers have remained at the forefront of astronomy and astrophysics during this period, maintaining a position of world leadership established early in this century. Spectacular advances in x-ray, gamma-ray, infrared, and ultraviolet astronomy

—supplementing the more established techniques of optical and radio astronomy—have opened virtually the entire electromagnetic spectrum to cosmic observations. The U.S. scientific community stands ready to exploit this opportunity with present and planned facilities of unparalleled power and scientific promise.

Special "decade reviews" carried out by the National Academy of Sciences (NAS) have periodically surveyed the status of the field and have advanced recommendations for future research programs. The first two such surveys, completed in 1964 and 1972, had a powerful impact on the conduct of U.S. research during the period 1960–1980. The recommendations of the third such decade review are contained in the recent report of the Astronomy Survey Committee, *Astronomy and Astrophysics for the 1980's* (National Academy Press), Washington, D.C., 1982; see Appendix A). Timely response to these recommendations will permit U.S. scientists to maintain leadership in the most promising areas of astronomy and astrophysics during the 1980's.

The Astronomy Survey Committee report took three years to complete and engaged over 130 scientists. Its final recommendations were highly selective: of many dozens of projects considered, only a small number were recommended for new funding for the 1980's (see Appendix B for a summary taken from the 1982 report). The National Academy of Sciences has accepted these recommendations as the consensus of the U.S. astronomical community, and they have also received the official endorsement of the American Astronomical Society.

The present Briefing Panel has based its recommendations for FY 1984 entirely on the recently completed Survey Committee report, and they follow the same format as does the report. The Survey Committee first stressed the importance of *Approved, Continuing, and Previously Recommended Programs* for progress in research during the 1980's. The Committee next emphasized that *Prerequisites for New Research Initiatives* are essential to the success of present and planned major research facilities. Finally, the Committee recommended the funding of *New Programs* for the 1980's in three categories: Major, Moderate, and Small. Throughout its study, the Survey Committee strove to be fiscally realistic: the overall program recommended for the 1980's is roughly comparable in scale with that actually carried out during the 1970's on the basis of the recommendations of the previous decade review.

From the projects recommended for the 1980's in the Survey

Committee report, the Briefing Panel has selected for recommendation here only those that are particularly timely for FY 1984 budget action. (It is important to recall again that the programs recommended by the Survey Committee themselves reflected a careful selection from a great many proposals made to that Committee.) Favorable action by the government in FY 1984 on the programs recommended here will constitute an important step toward ensuring the vitality of U.S. astronomical research in the decade ahead. The Panel calls particular attention to the need for FY 1984 commitments to begin two important projects: the Shuttle Infrared Telescope Facility (SIRTF) within the NASA Spacelab program and a Very Long Baseline Array (VLBA) of radio telescopes within the program of the NSF Astronomy Division.

II. THE RECOMMENDATIONS

The recommendations of the Briefing Panel for FY 1984 are summarized in this section in the same format as in the Survey Committee report. They are discussed in more detail in Section III.

APPROVED, CONTINUING, AND PREVIOUSLY RECOMMENDED PROGRAMS

Shuttle Infrared Telescope Facility (SIRTF). This major Spacelab facility is the only space-astronomy project in this category that has not yet been started. SIRTF will achieve thousandfold gains over present telescopes in the important thermal-infrared region of the spectrum. The Panel recommends that NASA re-establish its original strong commitment to a vigorous Spacelab program by immediately implementing SIRTF through a Phase B study, which will require an augmentation to the Spacelab program of $2 million in FY 1984 and $2.5 million in FY 1985.

PREREQUISITES FOR NEW RESEARCH INITIATIVES

Instrumentation and Detectors, Theory and Data Analysis, and Computational Facilities. The Survey Committee identified five Prerequisites which are essential to the success of major research programs but which are inexpensive by comparison. The Panel has selected the three Prerequisites stated above because funding them will immediately increase the efficiency with which observations are made and analyzed and the degree to which the resulting data can be understood on the basis of physical laws. The Panel recommends that an augmentation of $10 million/year above present funding levels, to be shared by NSF and NASA, be made available for these Prerequisites beginning in FY 1984.

NEW PROGRAMS

MAJOR NEW PROGRAMS

Advanced X-Ray Astrophysics Facility (AXAF)
This facility, which was accorded the highest priority among all major new programs by the Survey Committee, will be the nation's first permanent x-ray observatory in space. The Panel recommends augmentation of Phase B design studies by $2.5 million in FY 1984 and by $5 million in FY 1985 in order to ensure a new start for AXAF in FY 1986.

Very Long Baseline Array (VLBA) of Radio Telescopes
This ground-based facility, which was accorded the highest priority among major new ground-based programs by the Survey Committee (and second highest priority among all major new programs), will permit radio astronomers to map detail in cosmic radio sources with an angular resolution of 0.2 milliarcsecond. The Panel recommends that, in FY 1984, NSF commit to the construction of the VLBA and ensure its timely completion by making available $2–3 million for a final design study in FY 1984. Construction will require additional funding of $60 million spread over FY 1985–1987.

New Technology Telescope (NTT) of the 15-Meter Class
A ground-based NTT, which was accorded third highest priority among all major new programs by the Survey Committee, will make unique contributions to optical and infrared astronomy, particularly through optical and infrared spectroscopy of faint stars and galaxies and through high angular resolution studies in the infrared. The Panel recommends that design studies for the NTT, which were considered to be of the highest priority by the Survey Committee, be augmented to the level of $2 million in FY 1984 and in FY 1985 so that a final design for NTT can be selected by FY 1986.

MODERATE NEW PROGRAM

Augmentation to the NASA Explorer Program

The Explorer program remains a flexible and highly cost-effective means to pursue new opportunities in space astronomy in nearly every part of the electromagnetic spectrum. The Survey Committee recommended an immediate and substantial augmentation to the Explorer budget in order to restore it to the healthy level of effort of 1970, which prevailed before the ravages of inflation. The Panel recommends that the Explorer budget be augmented by $20 million/year beginning in FY 1984 in order to accelerate the development and launch of three Explorer missions now in various stages of preparation: the Cosmic Background Explorer (COBE), the Extreme Ultraviolet Explorer (EUVE), and the X-Ray Timing Explorer (XTE).

III. JUSTIFICATION FOR THE RECOMMENDATIONS

APPROVED, CONTINUING, AND PREVIOUSLY RECOMMENDED PROGRAMS

Shuttle Infrared Telescope Facility (SIRTF)

The programs recommended in the Astronomy Survey Committee's report were selected from research activities that were, at the beginning of the survey, candidates for implementation in FY 1983 and beyond. However, the Committee also emphasized "the importance of approved, continuing, and previously recommended programs to the progress of astronomical research during the remainder of the decade. The present Committee's recommendations take explicit account of such programs and build upon them" (Chapter 2, page 13; see also Appendix B of this Briefing Panel report). The Panel notes that SIRTF is the only space-astronomy project in this category that has not yet been implemented. The Survey Committee went on to say (Chapter 4, page 112—see also Appendix C to this Briefing Panel report), "The proposed SIRTF will be the cornerstone of research in infrared astronomy during the 1980's. The Astronomy Survey Committee joins with the Space Science Board's Committee on Space Astronomy and Astrophysics . . . in recommending this facility as the first major infrared telescope in space." The Panel enthusiastically supports this recommendation.

In support of its recommendation for SIRTF, the Survey Committee wrote: "SIRTF will permit investigations over the enormous range of wavelengths from 2 to 300 μm. For some important types of observations it will, because of its cryogenically cooled optics, yield a sensitivity gain of 1000 over the largest existing ground-based and airborne infrared telescopes; this gain in sensitivity is so large that it is not unreasonable to expect that SIRTF will make important and unexpected discoveries."

SIRTF will make it possible to observe the highly red-shifted light originating far in the past from very distant galaxies, so that we may study their early evolution. It will be able to detect very low-mass stars in nearby galaxies, which may account for the hidden mass in those galaxies, and to penetrate the obscuring dust of the dense cores of molecular clouds to study the process of star formation. SIRTF will also be able to investigate the thermal energy balance of the more distant planets and to greatly improve our knowledge of their atmospheric compositions.

Because of its wavelength range and sensitivity, SIRTF will in addition permit detailed spectroscopic studies of the infrared sources which will be discovered by the IRAS infrared survey satellite due to be launched in the very near future. Just as the 200-inch Hale telescope is used to investigate the stars and galaxies discovered by the 48-inch Schmidt telescope on Mt. Palomar, SIRTF will investigate new infrared sources discovered by IRAS.

A reaffirmation of NASA's original strong commitment to Space Shuttle research beginning in FY 1984 will be necessary to permit the timely development of SIRTF within the Spacelab program. With the flight of IRAS the development of cryogenically cooled optical systems for infrared space observations will have been completed and tested in space, thus removing a major barrier to the rapid development of SIRTF, a much more powerful and sensitive facility.

PREREQUISITES FOR NEW RESEARCH INITIATIVES

The Astronomy Survey Committee called attention to five Prerequisites for New Research Initiatives and recommended augmentations in each: "These Prerequisites are essential for the success of major research programs but are inexpensive by comparison." The Briefing Panel recommends that three of the Prerequisites receive augmentations in FY 1984.

Instrumentation and Detectors

Great strides in the development of new instruments and detectors were made during the 1970's, permitting enormous increases in observing efficiency, but so far only a handful of the telescopes at major observatories have been equipped with state-of-the-art instruments and detectors. Moreover, new generations of instruments and detectors must be developed in concert with the design and operation of future ground-based and space observatories. The application of state-of-the-art technology will greatly amplify the quantity and quality of the data obtained, at a small fraction of the cost of the major facilities that will use these instruments and detectors. Augmented funding for instrumentation will also permit accelerated development of a 10-meter submillimeter-wave radio antenna—the Astronomy Survey Committee's highest-priority recommendation for small new programs—by the NSF.

Theory and Data Analysis

The aim of astronomy is an understanding of the Universe, not simply the gathering of observational data. Vigorous programs of data analysis and complementary theoretical investigations are essential to achieve this aim, but both have been underfunded in the past. In particular, NASA has not adequately supported the theoretical analysis of data from space astronomy missions. Moreover, theoretical understanding of recent discoveries is important for optimizing the designs of future instruments. Increased support by NASA and NSF for theory and data analysis will greatly enhance the scientific return from both ground-based and space-astronomy facilities.

Computational Facilities

Modern instrumentation and detectors are highly dependent upon computers for control, data readout, and on-line processing. Because of the complexity of astronomical systems such as exploding stars and galaxies, theoretical investigations and data analysis cannot be carried out by purely analytic methods, and are therefore also dependent upon numerical calculations. The steady and rapid decrease in the cost of computing power opens up vast new opportunities for scientific advance. An augmentation to NSF and NASA funding of computational facilities will be a major step in achieving maximum effectiveness of new instruments and data-analysis programs.

NEW PROGRAMS

MAJOR NEW PROGRAMS

Advanced X-Ray Astrophysics Facility (AXAF)

The United States pioneered world leadership in x-ray astronomy, beginning with the 1962 discovery of extrasolar x-ray sources by means of a brief rocket flight, through the first all-sky survey of x-ray sources carried out by the UHURU satellite in 1970, to the flight of High Energy Astronomical Observatory (HEAO) satellites during the period 1977–1981. As a result, x-ray observations have now attained an importance in contemporary astronomy comparable to those in other wavelength regions. In FY 1986, following completion of Phase B studies, NASA should begin construction of the Advanced X-Ray Astrophysics Facility (AXAF), the nation's first permanent x-ray observatory in space. The Astronomy Survey Committee accorded this project highest priority of all major new programs for the 1980's.

With up to 100 times greater sensitivity than any previous x-ray mission, AXAF will provide U.S. astronomers with a powerful and long-lived new capability. Working in concert with the Very Large Array and the VLBA in the radio region, and with Space Telescope in the visible and ultraviolet spectral regions, AXAF will be an important part of an ensemble of major U.S. facilities covering much of the electromagnetic spectrum and capable of observing objects simultaneously at different wavelengths; this capability will greatly increase the scientific value of the data obtained.

AXAF will be used to study faint x-ray sources in the farthest reaches of our Galaxy as well as individual high-luminosity sources in hundreds of other galaxies. High-resolution imaging, together with spectroscopic and polarimetric observations, will reveal the composition and dynamics of supernova remnants, galactic halos, and clusters of galaxies. Observations of remote galaxies and quasars will probe the effects of evolution in the early Universe.

Augmentations of funding for final AXAF design studies in each of FY 1984 and FY 1985 are necessary for the timely completion of Phase B studies.

Very Long Baseline Array (VLBA) of Radio Telescopes

The United States has led the world in the development of Very Long Baseline Interferometry (VLBI), a technique that permits the combination in a central computer of radio-frequency observations from widely separated observing sites to produce images of cosmic objects with extraordinary angular resolution. The Very Long Baseline Array (VLBA) will include about 10 antennas, placed in Alaska and Hawaii as well as across the continental United States. This facility will map cosmic radio sources of very small angular size with a fineness of detail equal to that which could be achieved by a single radio dish as large as an entire continent. The angular resolution achieved—0.2 milliarcsecond—is equivalent to being able to read, from Los Angeles, a *Post* headline in Washington, D.C. Among many applications of profound importance, this instrument will probe the small-scale structure surrounding the enigmatic energy sources in the cores of quasars and active galactic nuclei and will directly determine the distance scale within our Galaxy with unprecedented accuracy (see Appendix D).

The National Science Foundation is currently completing a peer review of the VLBA, on the basis of which the Panel feels that NSF should be in a position to make a positive commitment to this important program. Accordingly, the Panel recommends that in FY 1984 NSF commit to construction of the VLBA and ensure its timely completion through award of $2–3 million for a final design study in FY 1984. The Array utilizes proven technology, and its costs may be accurately estimated. Construction of the VLBA will require additional funding of $60 million spread over FY 1985–1987. However, it will be critically important for NSF to construct and operate the VLBA without a further erosion of programmatic base funds within the NSF Astronomy Division.

Approval of the VLBA will mark the end of a long hiatus in the construction of new ground-based astronomical facilities by NSF. The 1970's saw the completion of the Very Large Array radio telescope near Socorro, N.M., but construction of a 25-meter millimeter-wave radio telescope—also recommended by the 1972 decade review committee—was deferred for a number of years for fiscal reasons and is not at present being considered for funding in the form originally proposed. During this period, other nations have taken over the lead from the United States in the important field of millimeter-wave studies of atoms and molecules in space. The field of VLBI, which was also pioneered in the United States, represents an outstanding opportunity for maintaining leadership in this exciting field of science.

New Technology Telescope (NTT) of the 15-Meter Class

In considering this project, the Astronomy Survey Committee wrote: "The Committee finds the scientific merit of this instrument to be as high as that of any other facility considered and emphasizes that its priority ranking (No. 3 among major new programs) does not reflect its scientific importance but rather its state of technological readiness. *The design studies needed before NTT can be constructed are of the highest priority and should be undertaken immediately.*" Thus, the Committee emphasized the importance of an immediate start toward the ultimate goal of NTT. The NTT has been studied at a low level for the past several years; because of the technological challenge of the NTT project, an augmentation of these studies to the level of $2 million/year in each of FY 1984 and 1985 will be required to permit a final design to be selected by FY 1986.

A ground-based telescope in the 15-meter class will make powerful contributions to optical and infrared astronomy, particularly through spectroscopic observations of faint stars and galaxies that cannot practicably be carried out with present instruments and through infrared observations with high angular resolution. New technology developed during the 1970's has made it feasible to build telescopes on this scale at a cost far lower than would have been possible only a few years ago. The capabilities of NTT will complement those of Space Telescope, which has higher angular resolution but less light-gathering power in the optical region of the spectrum, and of SIRTF, which has higher sensitivity but lower angular resolution in the infrared spectral range. The spectroscopic studies required to follow up on the discoveries by ST can be carried out only with an instrument having the aperture of NTT.

MODERATE NEW PROGRAM

Augmentation to the NASA Explorer Program

Among moderate new programs recommended for the 1980's, the Astronomy Survey Committee listed an augmentation to the NASA Explorer program as the highest priority: ". . . the flight of new Explorer missions has in recent years fallen much below the rate needed for healthy advance. The rate will decline even more drastically during the early 1980's if present budget levels are not increased. The Astronomy Survey Committee thus recommends an immediate and substantial augmentation to the Explorer program to restore it to at least the real level of effort of 1970." An augmentation of $20 million in FY 1984 represents an important first step in response to this recommendation. Such funding will permit the accelerated development and launch of three Explorer missions currently in various stages of preparation.

1. *The Cosmic Background Explorer (COBE),* now under development, will test the big-bang model of the Universe by extremely precise measurements of the cosmic microwave background radiation.
2. *The Extreme Ultraviolet Explorer (EUVE),* at present under study, will conduct the first all-sky survey of EUV sources, closing an important gap in our knowledge of the cosmic electromagnetic spectrum.
3. *The X-Ray Timing Explorer (XTE),* also now under study, will permit studies of the variability of x-ray sources on time scales from milliseconds to years, providing new constraints on models of neutron stars, accretion disks around collapsed stars, x-ray bursters, and quasars.

APPENDIX A

[REPRINTED FROM *SCIENCE,* AUGUST 20, 1982, VOLUME 217, NO. 4561, P. 691. COURTESY, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.]

ASTRONOMY AND ASTROPHYSICS IN THE UNITED STATES

A recently issued report* on astronomy begins: "Nature offers no greater splendor than the starry sky on a clear, dark night. Silent, timeless, jeweled with the constellations of ancient myth and legend, the night sky has inspired wonder throughout the ages."

For most of human history, leadership in studying the heavens has resided elsewhere, but during the 20th century the United States has been the world center of astronomy. This preeminence has been due to good financial support and the imaginative creation of innovative observing equipment. The capabilities of excellent optical telescopes, developed during the first half of this century, were later extended by equipment designed for observing throughout the electromagnetic spectrum. Leading supporters of the development of optical telescopes

*_Astronomy and Astrophysics for the 1980's,_ volume 1, *Report of the Astronomy Survey Committee* (National Academy Press, Washington, D.C., 1982).

were the Carnegie Institution of Washington, with its 2.5-meter telescope at Mount Wilson, and the Rockefeller Foundation, which gave the California Institute of Technology funds to build the 5-meter telescope at Mount Palomar. More recently, the National Science Foundation has become a major funder of ground-based astronomy, while NASA has provided excellent facilities in space. The United States has led in exploration of the solar system. In addition, it has launched space vehicles that have permitted observations which could not be achieved from the earth because of absorption of radiation in the atmosphere. The Space Telescope, to be launched in 1985, will be free from atmospheric inhomogeneities that blur sources of light and will be capable of high resolution of objects.

By 1970, generous support of American astronomy had led to many discoveries, including Hubble's expanding universe, time and celestial distance scales, quasars, x-ray sources, high-energy celestial gamma rays, the cosmic microwave background radiation, and polyatomic molecules in interstellar clouds. Discoveries during the 1970's included neutron stars accreting matter from nearby companion stars, hot intergalactic gas whose mass rivals that of the galaxies themselves, vast regions of interstellar gas heated to hundreds of thousands of degrees by shock waves from supernova explosions, and a gravitational lens effect observed as the splitting of light from a distant quasar as the light passed through an intervening galaxy.

The contributions of American astronomy are important and impressive. However, leadership cannot be maintained by resting on our laurels. Continuing preeminence of the United States will be dependent on well-trained people, who are provided with superior equipment.

The astronomical community has made a careful and searching study of opportunities and needs for support for the 1980's. Through extensive consultation and deliberation, a consensus has been achieved. The major new equipment recommended includes (i) an advanced X-ray Astrophysics Facility operated in space; (ii) a Very-Long-Baseline Array of radio telescopes; (iii) a New Technology Telescope, 15 meters in diameter, for ground-based studies in the optical and infrared regions of the spectrum; and (iv) a Large Deployable Reflector in space. All of these proposals would substantially extend the capabilities of astronomy. For example, the Very-Long-Baseline Array would have an angular resolution 100 times better than that of any other image-forming telescope at any wavelength. It would yield detailed radio images of quasars, the nuclei of galaxies, and features of interstellar molecular clouds and other astronomical objects. The first and third items above would be important for many studies, perhaps the most interesting being the examination of extremely distant objects whose radiation was emitted early in the history of the universe.

The report is well constructed and readable. It states well the case for additional expenditures for astronomy. Because of current budgetary problems, its recommendations may not be quickly accepted. However, it is designed to be relevant to the 1980's and at least part of it will surely be ultimately implemented.—PHILIP H. ABELSON

APPENDIX B

[REPRINTED FROM *ASTRONOMY AND ASTROPHYSICS FOR THE 1980'S, VOLUME 1, REPORT OF THE ASTRONOMY SURVEY COMMITTEE* (WASHINGTON, D.C.: NATIONAL ACADEMY PRESS, 1982), PP. 13–22.]

The Astronomy Survey Committee takes note at the outset of the support provided to U.S. astronomy and astrophysics over the past decades through the scientific programs of the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), and other federal agencies. This support has enabled U.S. astronomical research to maintain an overall position of world leadership and has vastly widened our horizons for exploration of the Universe.

The programs recommended in this report have been selected from research activities that were, at the beginning of the Survey, candidates for implementation in fiscal year 1983 and beyond. Before presenting a summary of its recommendations, however, the Committee wishes to emphasize the importance of approved, continuing, and previously recommended programs to the progress of astronomical research during the remainder of the decade. The present Committee's recommendations take explicit account of such programs and build upon them.

The Committee calls particular attention to the need for support of the following approved and continuing programs, for which the order of listing carries no implication of priority: Space Telescope and the associated Space Telescope Science Institute; second-generation Space Telescope instrumentation; the Gamma Ray Observatory; NASA level-of-effort observational programs, including research with balloons, aircraft, and sounding rockets, together with the Explorer and Spacelab programs; the Solar Optical Telescope and the Shuttle Infrared Telescope Facility for Spacelab; facilities for the detection of neutrinos from the solar interior; federal grants in support of basic astronomical research at U.S. universities; and programs at the National Astronomy Centers. The 25-Meter Millimeter-Wave Radio Telescope, which was recommended in an earlier form in the Greenstein report, has not yet been implemented. The present status of these approved, continuing, and previously recommended programs is described later in this chapter; their importance for the health of U.S. astronomy in the 1980's is discussed in Chapter 4.

SUMMARY OF THE RECOMMENDED PROGRAM

The Astronomy Survey Committee recommendations for a program in astronomy and astrophysics for the 1980's fall into three general categories:

Prerequisites for new research initiatives;
New programs; and
Programs for study and development.

As noted in the Preface, the observational components of these recommendations are restricted to remote sensing from the Earth or its vicinity. A background and overview of the recommendations follow later in this chapter.

Prerequisites for New Research Initiatives

In order to be effective, the recommended new research initiatives for the coming decade must be supported by a set of Prerequisites that apply to both the gathering and the analysis of the data produced. These Prerequisites are essential for the success of major programs but are inexpensive by comparison. Although significant support already exists for each, the Committee strongly recommends substantial augmentations in the following areas, in which the order of listing carries no implication of priority:

- **A.** *Instrumentation and detectors,* to utilize the latest technology to enhance the efficiency of both new and existing telescopes in the most cost-effective manner;
- **B.** *Theory and data analysis,* to facilitate the rapid analysis and understanding of observational data;
- **C.** *Computational facilities,* to promote data reduction, image processing, and theoretical calculations;

D. *Laboratory astrophysics*, to furnish the atomic, molecular, and nuclear data essential to the interpretation of nearly all astronomical observations; and

E. *Technical support at ground-based observatories*, to ensure that modern astronomical instrumentation is maintained in the best condition permitted by the state of the art.

A detailed consideration and justification of these Research Prerequisites appears in Chapter 5.

New Programs

The Astronomy Survey Committee recommends the approval and funding of new programs in astronomy and astrophysics for the 1980's. These have been arranged into three categories according to the scale of resources required.

A. *Major New Programs* The Committee believes that four major programs are critically important for the rapid and effective progress of astronomical research in the 1980's and is unanimous in recommending the following order of priority:

1. *An Advanced X-Ray Astrophysics Facility* (AXAF) operated as a permanent national observatory in space, to provide x-ray pictures of the Universe comparable in depth and detail with those of the most advanced optical and radio telescopes. Continuing the remarkable development of x-ray technology applied to astronomy during the 1970's, this facility will combine greatly improved angular and spectral resolution with a sensitivity up to one hundred times greater than that of any previous x-ray mission.
2. *A Very-Long-Baseline* (VLB) *Array* of radio telescopes designed to produce radio images with an angular resolution of 0.3 milliarcsecond. Among many potential applications of profound importance, this instrument will probe the small-scale structure surrounding the enigmatic energy sources in the cores of quasars and active galactic nuclei and will directly determine the distance scale within our Galaxy with unprecedented accuracy.
3. *A New Technology Telescope* (NTT) of the 15-m class operating from the ground at wavelengths of 0.3 to 20 μm, to provide a tenfold increase in light-gathering capacity at visual wavelengths and a hundredfold increase in speed for spectroscopy at infrared wavelengths, with application to a very wide range of scientific

problems. The Committee finds the scientific merit of this instrument to be as high as that of any other facility considered and emphasizes that its priority ranking does not reflect its scientific importance but rather its state of technological readiness. *The design studies needed before* NTT *can be constructed are of the highest priority and should be undertaken immediately.*

4. *A Large Deployable Reflector* in space, to carry out spectroscopic and imaging observations in the far-infrared and submillimeter wavelength regions of the spectrum that are inaccessible to study from the ground, thus extending the powerful capabilities of NTT to these longer wavelengths. Such an instrument, in the 10-m class, will present unprecedented opportunities for studying molecular and atomic processes that accompany the formation of stars and planetary systems.

B. *Moderate New Programs* In rough order of priority, these are:

1. *An augmentation to the* NASA *Explorer program,* which remains a flexible and highly cost-effective means to pursue important new space-science opportunities covering a wide range of objects and nearly every region of the electromagnetic spectrum.
2. *A far-ultraviolet spectrograph in space,* to carry out a thorough study of the 900-1200-Å region of the spectrum, important for studies of stellar evolution, the interstellar medium, and planetary atmospheres.
3. *A space* VLB *interferometry antenna* in low-Earth orbit, to extend the powerful VLBI technique into space in parallel with the rapid completion of a ground-based VLB Array, in order to provide more detailed radio maps of complex sources, greater sky coverage, and higher time resolution than the Array can provide alone.
4. *The construction of optical/infrared telescopes in the 2–5-m class,* to observe transient phenomena, conduct long-term survey and surveillance programs, provide crucially needed ground-based support to space astronomy, and permit the development of instrumentation under realistic observing conditions. The Committee particularly encourages federal assistance for those projects that will also receive significant nonfederal funding for construction and operation.
5. *An Advanced Solar Observatory* in space, to provide observations of our Sun—the nearest star—simultaneously at optical, extreme ultraviolet, gamma-ray, and x-ray wavelengths, to carry out long-

term studies of large-scale circulation, internal dynamics, high-energy transient phenomena, and coronal evolution.

6. *A series of cosmic-ray experiments* in space, to promote the study of solar and stellar activity, the interstellar medium, the origin of the elements, and violent solar and cosmic processes.
7. *An astronomical Search for Extraterrestrial Intelligence* (SETI), supported at a modest level, undertaken as a long-term effort rather than as a short-term project, and open to the participation of the general scientific community.

C. *Small New Programs* The program of highest priority is:

- An antenna approximately 10 m in diameter for submillimeter-wave observations, at an excellent ground-based site.

Other programs of outstanding scientific merit, in which the order of listing carries no implication of priority, are as follows:

- A spatial interferometer for observations of high angular resolution in the mid-infrared region of the spectrum;
- A program of high-precision optical astrometry; and
- A temporary program to maintain scientific expertise at U.S. universities during the 1980's through a series of competitive awards to young astronomers.

Detailed discussion and justification of the New Programs appears in Chapter 6.

Programs for Study and Development

Planning and development are often time-consuming, especially for large projects. It is therefore important during the coming decade to begin study and development of programs that appear to have exceptional promise for the 1990's and beyond. Projects and study areas recommended by the Committee in this category include the following, in which the order of listing carries no implication of priority:

A. Future x-ray observatories in space;
B. Instruments for the detection of gravitational waves from astronomical objects;
C. Long-duration spaceflights of infrared telescopes cooled to cryogenic temperatures;

D. A very large telescope in space for optical, ultraviolet, and near-infrared observations;
E. A program of advanced interferometry in the radio, infrared, and optical spectral regions;
F. Advanced gamma-ray experiments; and
G. Astronomical observatories on the Moon.

Detailed discussion of the Programs for Study and Development appears in Chapter 7.

ESTIMATED COST OF THE RECOMMENDATIONS

In order to establish the overall scale of the recommended total program, the Committee gives in Table 2.1 its own approximate estimates of the requirements for new funding over the next 10 years in millions of 1980 dollars. Funds for projects to be supported by NASA represent research-and-development funds within NASA's Office of Space Science and Applications (OSSA); funds for projects to be supported by NSF represent total cost to NSF. Operating costs are included for those facilities expected to become operational in the 1980's.

The funding entries for the Prerequisites for New Research Initiatives represent augmentations to the present levels of support for these activities within NSF and NASA. As it is expected that the two agencies will work together to coordinate support for the Prerequisites, specific agency responsibility is not indicated in the following table. However, since the Prerequisites provide support to space- and ground-based research at comparable levels, the Committee anticipates that the funding augmentations to be provided by NASA and NSF will be roughly equal in magnitude.

In the cases of the New Programs, the division between space- and ground-based projects is clear. Funds listed for the Explorer program represent an augmentation to NASA's level-of-effort budget for that program; the operations costs listed for ground-based projects, together with the temporary program to maintain scientific expertise at U.S. universities, represent further augmentations to the operations budget of NSF's Astronomy Division. Remaining New Program costs represent new-funding requirements for either NASA (new starts within OSSA) or NSF (major construction within the Astronomy Division).

The cost estimates for AXAF and the VLB Array were derived with the help of individual scientists participating in current studies and are based on reasonably complete rough designs. The actual cost of NTT, however, cannot be estimated until further studies indicate which of

TABLE 2.1
REQUIREMENTS FOR NEW FUNDING

	Prerequisites for New Research Initiatives	(Millions of 1980 Dollars)
A.	Instrumentation and detectors (doubling of present $15 million/year level of effort by increments over 10 years)	$ 75
B.	Theory and data analysis (augmentation by $5 million/year)	50
C.	Computational facilities (30 minicomputer systems installed at a rate of 5 systems/year, including operations)	20
D.	Laboratory astrophysics (augmentation by $2.5 million/year)	25
E.	Technical support at ground-based observatories, including 40 new support positions	20
	DECADE TOTAL PREREQUISITES	$ 190
	New Programs	
A.	*Major New Programs* In order of priority:	
	1. Advanced X-Ray Astrophysics Facility (AXAF)	$ 500
	2. Very-Long-Baseline (VLB) Array (including $15 million for operations)	50
	3. New Technology Telescope (NTT)	100
	4. Large Deployable Reflector in space	300
	Decade Subtotal	$ 950
B.	*Moderate New Programs* In rough order of priority:	
	1. Augmentation to Explorer satellite program	$ 200
	2. Far-ultraviolet spectrograph in space	150
	3. Space VLB interferometry antenna	60
	4. Optical/infrared telescopes in the 2–5-m class	20
	5. Advanced Solar Observatory in space	200
	6. Cosmic-ray experiments	100
	7. An astronomical Search for Extraterrestrial Intelligence (SETI)	20
	Decade Subtotal	$ 750
C.	*Small New Programs* Of highest priority:	
	■ 10-m submillimeter-wave radio antenna (including $2 million for operations)	$ 4
	Other important programs:	
	■ Spatial interferometer for the mid-infrared (including $1 million for operations)	3
	■ High-precision optical astrometry program	3
	■ Temporary program to maintain scientific expertise at U.S. universities	10
	Decade Subtotal	$ 20
	DECADE TOTAL, NEW PROGRAMS	$1,720

several alternative conceptual designs will be most cost-effective; the figure given in Table 2.1 is meant as a limit that the Committee recommends should not be substantially exceeded. The estimated cost of the Large Deployable Reflector in space is highly uncertain because instrumentation of this type has not yet been developed and launched. Most of the costs estimated for the Moderate New Programs should be reasonable approximations, as they are based on experience with previous instruments of a similar nature. The costs given for the augmentation to the NASA Explorer program and for SETI, however, should be regarded as target figures for the level of effort the Committee finds appropriate.

The total cost in new funding estimated for the Prerequisites and New Programs together is about $1.9 billion in 1980 dollars. By comparison, the Greenstein report (1972) recommended new programs with an estimated cost of $844 million in 1970 dollars, or approximately $1.7 billion in 1980 dollars, and most of those programs were in fact implemented. *The program recommended here for the 1980's is thus roughly comparable in scale with that actually carried out during the 1970's on the basis of the recommendations of the Greenstein report.*

The Committee wishes to emphasize, however, that the present recommendations will require substantial increases in the budget of the Astronomy Division of NSF, the agency primarily responsible for the support of ground-based astronomy. If, as anticipated, NSF will provide roughly half of the additional funds required for the Prerequisites for New Research Initiatives, an increase of about 30 percent in the Astronomy Division's operations budget over the real level of expenditures during the 1970's will be required for NSF to carry out its share of the recommended program over the next decade. Funds needed by NSF for major construction over the next 10 years will also be substantially higher than those expended during the 1970's, which saw the completion of only one major project, the Very Large Array, at a cost of $78 million. *The Astronomy Survey Committee believes that these increases in the NSF budget for ground-based astronomy are essential to maintain an effective partnership with space astronomy during the 1980's.*

BACKGROUND AND OVERVIEW

The Greenstein Report

The publication of *Astronomy and Astrophysics for the 1970's* (the Greenstein report) by the National Academy of Sciences in 1972 had a powerful impact on the development of U.S. astronomy and astrophysics during 1972–1982. The federal government on the whole responded positively to its recommendations, with the result that the facilities available to U.S. astronomers have enabled them to remain at the frontiers of research. Here we review the responses to the recommendations of that report and their impact on the progress of science.

Radio Astronomy and the VLA The highest-priority recommendation of the Greenstein report was the construction of a Very Large Array (VLA) radio telescope, together with increased support for smaller facilities. Funded by NSF, the VLA was constructed in stages during the 1970's and was formally dedicated in 1980; by far the largest and most complex ground-based astronomical facility established to date, the VLA was completed on schedule and within budget. VLA studies of radio sources are already having a large impact on both Galactic and extragalactic astronomy (see, for example, the cover of this report).

The recommended increase in funding for smaller radio-astronomical facilities did not materialize, however, nor has funding yet been provided for a recommended millimeter-wave radio telescope, then projected to have a diameter of 65 m and to be operable at wave-lengths down to 3 mm. Since the publication of the Greenstein report, the study of interstellar molecules at millimeter wavelengths has yielded insight into the process of star formation; as the science has progressively moved to shorter wavelengths, there has now evolved a need for a smaller, more precisely figured telescope of 25-m diameter, still offering high sensitivity and spatial resolution but operable at wavelengths down to 1 mm. The recommendation for a large centimeter-wave antenna was not implemented, although existing facilities for observations at wavelengths longer than 1 cm have been maintained and in some cases upgraded.

Optical Ground-Based Astronomy The second-priority recommendation was for a variety of steps to enhance the capability available to U.S. optical astronomers. A key proposal was the development and construction of a multiple-mirror telescope (MMT) with aperture equivalent to that of a conventional telescope in the 3.8–5.0-m range; this project was to be followed by the construction of a larger MMT, of 10–15-m aperture, if that proved feasible, or by a conventional telescope of 5-m aperture if it did not. A 4.5-m MMT has in fact been developed jointly by the Smithsonian Institution and the University of Arizona, becoming operational in 1979. However, neither a larger MMT nor a conventional 5-m telescope has been funded, with the result that the largest instrument available to U.S. optical astronomers is still the 5-m Hale telescope on Mt. Palomar, which went into operation 35 years ago. As a result, optical spectroscopy of the faintest galaxies and quasars discovered by radio and x-ray astronomers has not kept pace with new discoveries, even though these extremely distant objects are of great interest because of their bearing on the nature of cosmic evolutionary processes early in the history of the Universe.

The recommendation also called for equipping existing telescopes with advanced electronic detectors and controls. Although major progress was made in the development of such devices during the 1970's, they have so far been provided to only a few major observatories. The capability of instruments now available for use on most of the nation's optical telescopes still lags far behind state-of-the-art technology.

The Greenstein report furthermore recommended that three telescopes in the 2.5-m class be constructed for a variety of purposes. As none of these has been funded, all the nation's major optical telescopes are heavily oversubscribed.

Infrared Astronomy The third-priority recommendation called for an across-the-board increase in support for infrared astronomy, which at that time was beginning to demonstrate its great importance. Support has increased substantially, through the funding of two major ground-based infrared telescopes (the 2.3-m University of Wyoming Infrared Observatory reflector and the 3-m Infrared Telescope Facility operated by NASA on Mauna Kea), the Kuiper Airborne Observatory program, and a balloon program. An international Infrared Astronomy Satellite (IRAS), scheduled for launch in 1982, will carry out a comprehensive, far-infrared survey of the sky, as called for in the Greenstein report.

APPENDIX C

[REPRINTED FROM *ASTRONOMY AND ASTROPHYSICS FOR THE 1980'S, VOLUME 1, REPORT OF THE ASTRONOMY SURVEY COMMITTEE* (WASHINGTON, D.C.: NATIONAL ACADEMY PRESS, 1982), PP. 112–113.]

THE SHUTTLE INFRARED TELESCOPE FACILITY (SIRTF)

The proposed SIRTF will be the cornerstone of research in infrared astronomy during the 1980's. The Astronomy Survey Committee joins with the Space Science Board's Committee on Space Astronomy and Astrophysics (*A Strategy for Space Astronomy and Astrophysics for the 1980's,* National Academy of Sciences, Washington, D.C., 1979) in recommending this facility as the first major infrared telescope in space.

SIRTF will permit investigations over the enormous range of wavelengths from 2 to 300 μm. For some important types of observations it will, because of its cryogenically cooled optics, yield a sensitivity gain of 1000 over the largest existing ground-based and airborne infrared telescopes; this gain in sensitivity is so large that it is not unreasonable to expect that SIRTF will make important and unexpected discoveries. The multiple, interchangeable focal-plane instruments planned for SIRTF will moreover greatly increase our ability to explore the evolution of distant extragalactic sources, the physical properties and chemical composition of molecular clouds and regions of star formation, the nature of cometary nuclei and asteroids, and the structure of planetary atmospheres. For example, SIRTF will be able to detect infrared sources at the limits of the observable Universe; on one Shuttle flight, it could gather information on sources of both large and small red shift, thus permitting a comparison of the energetics of quasars and galaxies at the earliest epochs of the Universe with those at the present epoch. Because of its relatively wide field of view, SIRTF will also be able to carry out efficient surveys of infrared sources that will help to optimize the observing programs of larger instruments, such as the New Technology Telescope (NTT), the VLB Array, and a Large Deployable Reflector (LDR) in space, which have narrower fields of view.

SIRTF should be an early and frequently flown payload on Shuttle sortie missions. It has such high sensitivity that very extensive infrared observations can be accomplished even within the relatively brief, 7-day profile of an early Shuttle flight; reflight on the Shuttle makes it possible to fly SIRTF with continually improved focal-plane instrumentation and detectors. Eventually, however, SIRTF flights of longer duration will be needed to realize the full potential of this remarkable facility, and the study of such flights is recommended in Chapter 7.

APPENDIX D

[REPRINTED FROM *ASTRONOMY AND ASTROPHYSICS FOR THE 1980'S, VOLUME 1, REPORT OF THE ASTRONOMY SURVEY COMMITTEE* (WASHINGTON, D.C.: NATIONAL ACADEMY PRESS, 1982), PP. 135–136.]

A VERY-LONG-BASELINE (VLB) ARRAY OF RADIO TELESCOPES

The Astronomy Survey Committee recommends the construction of a ground-based Very-Long-Baseline (VLB) Array of radio telescopes designed to produce images with an angular resolution of 0.3 milliarcsecond. Because the Array utilizes proven technology, this project may be begun immediately after completion of final management and design studies.

Extraordinarily high angular resolution is now possible at radio frequencies. Precision atomic clocks, more sensitive and reliable receivers, high-speed tape recorders, sophisticated image-processing techniques, and modern antennas now make it feasible to build a radio array with the angular resolution of a telescope covering an entire continent. This may be done by synchronizing the operation of about ten widely spaced antennas of approximately 25-m diameter, whose outputs are recorded and later combined in a central computer.

This VLB Array will produce high-quality radio images capable of resolving features down to 0.3 milliarcsecond (the size of a dime in New York as seen from Los Angeles). This is a hundred times better angular resolution than that of any other image-forming telescope at any wavelength and will yield detailed new radio images of a wide range of astronomical objects at the frontiers of modern astrophysical research. These include quasars and the nuclei of galaxies, features of interstellar molecular clouds, the centre of our Galaxy, and a variety of energetic Galactic objects such as x-ray, binary, and flare stars. The high angular resolution of the VLB Array will permit the direct study of small-scale structure surrounding the central regions of quasars and stars in the process of formation. Through the method of statistical parallaxes, it will furthermore permit direct measurements of distances to many objects throughout our Galaxy and even to some in nearby galaxies. The VLB Array can also be applied to important problems in Earth science (including precision geodesy and geophysics), to the navigation of interplanetary spacecraft, and to tests of the General Theory of Relativity.

Although the VLB Array is a complex and sophisticated instrument, it will make use of proven concepts and instrumentation. Construction should begin immediately upon completion of management and design studies with the building of the antennas and the development of the data-reduction system and other instrumentation. Collaboration with groups in other countries, particularly in Europe and North America, would improve the performance of the instrument by increasing the resolution even further (particularly in the north-south direction) and by improving the image quality at low declinations.

BRIEFING PANEL ON AGRICULTURAL RESEARCH

Lawrence Bogorad ***(Co-chairman),*** *Harvard University*

William L. Brown ***(Co-chairman),*** *Pioneer Hi-Bred International, Inc.*

Myron K. Brakke, *USDA-ARS, University of Nebraska*

Vincent G. Dethier, *University of Massachusetts*

Michael Freeling, *University of California, Berkeley*

Ralph W. F. Hardy, *E.I. duPont de Nemours & Company, Inc.*

Gary H. Heichel, *USDA-ARS, University of Minnesota*

Arthur Kelman, *University of Wisconsin*

David W. Krogmann, *Purdue University*

Carole P. Meredith, *University of California, Davis*

Oliver E. Nelson, Jr., *University of Wisconsin*

Staff

James E. Tavares, ***Senior Staff Officer,*** *Board on Agriculture*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON AGRICULTURAL RESEARCH

SUMMARY

This report focuses almost exclusively on plant science research related to improved agricultural production efficiency. An assessment is made of the realistic expectations of genetic engineering applications for improving crops and cropping practices; several promising research areas for increased federal research investment are identified; and some constraints to plant science research are briefly discussed.

GENETIC ENGINEERING

In the near future the biggest contribution of the application of plant molecular biology and genetic engineering techniques to plants will be to fundamental knowledge. Such new knowledge in plant science will have enormous potential for making major contributions to crop improvement and cropping practices.

Genetic engineering has already provided unexpected new insights into the organization and function of genetic traits. The direct demonstration of relationships between regulatory and structural genes, the presence of repeated sequences of DNA throughout the genome, the existence of introns or noncoding sequences within genes, and the surprising plasticity of the genome that was earlier thought to be rigidly conservative have all helped and forced scientists to think in new ways about the factors that influence and control the reassortment and expression of hereditary traits. Much of this increased understanding could only be derived through the use and analysis of the technologies associated with genetic engineering.

Genetic engineering is a powerful adjunct to plant breeding. The successful application of genetic engineering will generate new and previously unavailable sources of desirable genetic variability that can be exploited through conventional plant-breeding techniques. The application of cell culture techniques is already contributing new sources of genetic variation to plant breeding.

While the application of gene-splicing techniques in plants is still in its infancy, it seems reasonable to expect that there are traits controlled by a single gene or a small group of genes (such as for herbicide resistance) that can be transferred successfully using genetic engineering. Important questions remain, however, on whether the introduction of such foreign DNA will affect the expression of other genes in the plant and whether coordinated multigene traits can be engineered successfully.

RESEARCH OPPORTUNITIES

Advances in molecular biology techniques and improved instrumentation are providing new opportunities in plant science research. In the mid-1970s several high-priority research areas in the plant sciences were identified, and four of these areas—photosynthetic enhancement, biological nitrogen fixation, genetic mechanisms, and plant biological stress—received further recognition through increases in federal financial support. The resulting expansion of research programs has been highly productive.

The following seven research opportunity areas were identified by the Briefing Panel on Agricultural Research as holding particular promise for establishing the basic scientific understanding that could lead to important discoveries and applications for crop improvement. These areas are listed in order of priority.

Genome Structure and Gene Expression in Plants

Basic to all plant science research is an understanding of the genetic information system that controls plant growth and function. Essential to progress in these areas are the identification and isolation of specific genes in plants, the development of techniques for inserting the DNA of other plants and organisms into plant cells, and the establishment of *in vitro* systems that will permit the genetic expression of such foreign DNA.

Plant Developmental Biology

Discoveries in the regulation of gene expression in plants will greatly aid in an understanding of the control and development of form and function in plants. Conversely, the elucidation of factors involved in intercellular communication and the coordination of processes on an organismic level in plants should provide important insights for studying the genetic controls of plant development.

Plant and Pest Interactions

The genetic and biochemical mechanisms that control virulence in plant pathogens and the related resistance that can be established in crop cultivars are now amenable to analysis because of the availability of the powerful tools of molecular biology. The wealth of information on genome structure and the control of gene expression in viruses and bacteria can hasten advances in the understanding of the actions of these and other pathogens. Continued and accelerated work in identifying bacterial toxins and host-specific fungal toxins is needed, along with work on the action of chemicals that trigger systemic defense mechanisms in host plants. Multidisciplinary studies combining insect neurobiology and the analytical chemistry of secondary plant metabolites that attract or repel insects should lead to useful applications for controlling pest insects.

Organisms in the Rhizosphere

Little is known about the ecology of the organisms that inhabit the root zone. The intriguing properties of disease-suppressive soils and the growing recognition of the importance of the ubiquitous mycorrhizal fungi point to the enormous potential benefits that could be derived from intensive study of these soil organisms.

Photosynthetic Energy Conversion and Carbon Metabolism

This report discusses in some detail the considerable progress that has been made in understanding the components of photosynthetic energy conversion. The methods and research tools are at hand to continue and expand this progress. In addition, increased emphasis needs to be placed on research in carbon metabolism, particularly on the control of partitioning the carbohydrate products of photosynthesis throughout the plant and on the feedback effects of this partitioning on net photosynthesis.

Mineral Uptake, Deficiencies, and Excesses
The existence of crop cultivars with impaired or improved uptake of essential mineral nutrients or resistance to toxic levels of other minerals provides opportunities for applying the tools of molecular biology to better understand mineral nutrition in plants. The benefits to be derived are twofold, namely, the more efficient production of crops on soils currently supplemented with high-nutrient inputs (particularly nitrogen) via a reduction in inputs, and the extension of cropping to marginal soils that are saline or contain toxic minerals.

Biological Nitrogen Fixation
Advances in the understanding of biological nitrogen fixation continue at all levels of research, from sequencing the nucleic acids in the genes of nitrogen-fixation systems to practical applications such as selecting or devising improved strains of nitrogen-fixing bacteria in their associations with crop plants. Many intriguing problems remain, such as why legumes do not show any yield response to nitrogen fertilizer as other crops do, how to control the wasteful loss of reducing power in the form of hydrogen gas generated by the nitrogen-fixation system, and how to achieve the ultimate goal of transferring functioning nitrogen-fixation genes directly into crop plants.

RESEARCH CONSTRAINTS

To realize the full potential of the research opportunities identified in this report, some attention also should be paid to factors, in addition to adequate funding, that influence the rate of progress in plant science research. These factors include the need for trained scientists and for institutional arrangements to encourage multidisciplinary research and to foster innovative scouting research.

Currently there is a shortage of scientists with training in both the basic plant sciences and agronomy as well as molecular biology and molecular genetics. This deficiency might be addressed in part through postdoctoral research opportunities to permit recent graduates in either plant science fields or molecular genetics to obtain research experience in the other area. At present there are relatively few postdoctorals in the agricultural sciences. Recent data published by the National Science Foundation show that there were 23,420 postdoctorals supported by federal sources in biological sciences, and 15,845 in health sciences, but only 821 postdoctorals in agricultural sciences for the period 1974 through 1979.* An alternative method for stimulating the interchange of ideas among disciplines and fostering cooperative research might be through special summer study programs. Two examples that might serve as models are the Cold Spring Harbor schools in molecular biology and molecular genetics and the Woods Hole Marine Biological Laboratory summer programs in several areas of biology. Both examples have been instrumental in advancing research in their areas.

There are perhaps additional ideas on how interdisciplinary research might be fostered, including the encouragement of multidisciplinary research teams and the opportunities to support the exploratory or scouting research that provides the foundation on which new, focused research programs and potential technologies are built.

INTRODUCTION

U.S. agriculture is continually changing. In contrast to the kinds of changes that occurred in earlier periods of the nation's history, the pace and nature of changes in agriculture today are to a considerable extent the result of scientific research and development. Such changes have helped account for this country's ability to supply itself with abundant, high-quality food and fiber at low cost and to help meet some of the food needs of the rest of the world.

Notable examples of wide-ranging effects of past research that have helped secure the success of the American agricultural system include the following: Fertilizers now permit the bountiful harvests that virgin soils were once capable of producing. Agricultural chemicals protect crops from insects and pathogens and have largely replaced mechanical cultivation practices that once controlled weeds. These same chemicals have found new uses in no-till or minimum-till farming that helps reduce one of the major problems of modern agriculture—soil erosion. Modern plant-breeding technology, based on a strong foundation of plant genetics, has produced crop varieties that are well adapted to particular environments. Plant breeders have kept ahead of

*National Science Foundation. 1981. *National Patterns of Science and Technology Resources.* NSF 81-131 (Washington, D.C.: U.S. Government Printing Office).

continuously evolving pest populations with a steady stream of resistant varieties.

Obviously, as with any type of scientific progress, these results in American agriculture were not brought about overnight by single-handed effort or by a few isolated discoveries. The results have come about in the same way that any organism grows—through a vast number of interrelated developments. The sustained success of American agriculture depends upon the continuation of research programs attuned to the problems of the present and to the changes needed in the future.

This report focuses almost exclusively on plant science research, the chief beneficiary of which traditionally has been agriculture. The report summarizes the discussions of a panel of 11 members who were asked to identify promising research opportunities in the plant sciences that could lead to developing more productive disease-resistant, stress-tolerant, and energy-efficient crop plants and to provide some assessment of the realistic expectations of genetic engineering for developing such crops. During three days of discussion several areas were identified as being particularly promising opportunities for highly effective use of incremental increases in federal research investment. These research areas include the following: genome structure and gene expression in plants; plant developmental biology; plant and pest interactions; organisms in the rhizosphere; photosynthetic energy conversion and carbon metabolism; mineral uptake, deficiencies, and excesses; and biological nitrogen fixation.

RESEARCH OPPORTUNITIES

GENOME STRUCTURE AND GENE EXPRESSION IN PLANTS

In a crop plant there are approximately 100,000 genes in each nuclear genome, that is, the complete set of genes found in the nucleus of each plant cell. At its simplest level, each gene is a DNA (deoxyribonucleic acid) sequence designed along similar lines, and each involves a unit of information used to specify a product called messenger RNA (ribonucleic acid), which in turn dictates the structure of proteins that control the plant's functions. Little is known about the molecular structure and organization of the genome or about factors regulating the expression of the individual genes. This is true of animals as well as of plants. An understanding of genome structure and gene function in plants is growing, however, because the means of acquiring answers to many questions are now available. The new tools of molecular biology and genetic engineering that offer prospects of identifying, isolating, and cloning genes will permit an understanding of the control of gene expression with a view to transferring such genes into crop plants.

The generalized function of some of the DNA sequences within a gene is known. There are DNA coding sequences that specify the amino acid sequence of proteins, and DNA promoter sequences that may bind enzyme molecules that initiate or prevent the initiation of the transcription of the gene into RNA. Stretches of DNA, called introns, reside within the coding sequences of most nuclear genes and are present in the RNA first transcribed from the DNA sequence, but are removed from this primary RNA as it becomes the messenger RNA that is translated into protein. The function of introns is not known. There are DNA sequences within or next to coding sequences that are thought to bind molecules that regulate the tissue-specific timing of gene expression during development. Also, there is evidence that changes in the organization or arrangement of genes in the genome can influence gene function. These levels of gene regulation are not understood. Fewer than 10 plant genes have been fully described as DNA sequences.

Recent advances in molecular biology include recombinant-DNA technology and rapid DNA-sequencing methods. In addition, transposable elements—DNA sequences that have the ability to move from place to place within the genome and affect the functioning of genes into which they insert—can be used as intragenomic probes to identify and isolate genes they associate with and are among the powerful techniques now available for gaining an understanding of the structure and regulation of genes in plants. Transposons, comparable mobile DNA sequences in bacteria, have been used successfully as such probes.

Specific areas requiring research include identification of which genes code for particular enzymes in plants and what molecules elicit the activity of such genes in a certain tissue or at a particular stage of development. An important extension of this knowledge would be the determination of what gene or genes could be transferred from one species to another with a view to improving crop productivity. This determination also requires the identification and understanding in molecular terms of the rate-limiting steps in plant growth and metabolism.

Knowledge of specific gene expression systems may provide information leading to control of rate-limiting steps. An example is the regulation of the bacterial nitrogen-fixation gene system by the chemical methionine sulfoxamine (MS). In the presence of high levels of ammonia or fixed nitrogen, bacterial nitrogen fixation will not occur. However, added MS causes the nitrogen-fixation gene system to be expressed even in the presence of quantities of fixed nitrogen that otherwise prevent expression of the system. This effect of MS provides an important example of how an understanding of the mechanism that regulates the expression of a biosynthetic system might lead to the identification and synthesis of specific growth-regulatory products that promote or arrest the system.

Plant growth regulators for increasing yield and improving crop quality have been sought for some time, but the success in identifying regulators for major agronomic crops has been very low. An understanding of plant gene structure and of gene expression systems should provide the essential knowledge base for identifying and synthesizing new plant growth regulators. Many suggest that the benefits of such as-yet-undiscovered plant growth regulators to agriculture may equal or exceed the benefits of plant protectants.

A particularly important area within the field of gene organization and expression involves the development of systems in which cloned DNA sequences may be assayed for function. The primary function of the DNA coding sequence of the gene involves the formation of an RNA intermediate, but this function for many genes is evoked only at particular times in development and often only in particular cells of the organism. There is specific need at present for the development of additional *in vitro* systems that permit the transcription of DNA coding sequences to RNA and their further translation to protein-enzyme products. More importantly, although astounding progress has been made in developing the crown gall bacterial plasmid system for plant genetic engineering, there is need for additional *in vivo* transformation systems in which genes can be integrated into the chromosomes of a cell, and that cell can, in turn, be used to generate an entire plant within which the gene's expression may be assayed. The cell wall can be removed from certain kinds of plant cells to produce protoplasts, and in some species these protoplasts can be regenerated into whole plants. Current research involving the introduction of foreign DNA into protoplasts is particularly promising.

In addition to the 100,000 or so nuclear genes in each plant cell, two plant cell organelles also contain genes. The cell's chloroplasts—the seat of photosynthesis—each contain perhaps 100 unique genes;

and mitochondria—the sites for oxidation of foodstuffs in the cell—also contain many genes. These three genetic systems of plant cells are highly interdependent, although features of their gene expression systems differ. The health, vigor, and success of a crop variety depend on organelle as well as nuclear genomes and on the proper integration of all three genetic systems. Furthermore, each of these genomes may have its own advantages as a site for introduction of genes for crop improvement.

A few laboratories in the United States currently are intensively investigating DNA sequences and their functions in plants. The knowledge base that these laboratories are developing is essential to the application of this technology to crop improvement.

PLANT DEVELOPMENTAL BIOLOGY

Plant developmental biology is the study of how a single cell gives rise to an entire plant with all its different tissues, organs, and functions. Much of the work in this area has been descriptive, but with the emergence of more sophisticated techniques, an understanding of molecular mechanisms of plant development is now possible. New and powerful tools of cellular and molecular biology promise to provide the means for achieving a much deeper understanding of the ultimate basis for the development of form and function in plants.

The identification and study of those genes and regulatory elements in the plant genome that have profound effects on development will lead to important advances in current knowledge of how the many genes in each cell are coordinately expressed to produce a complex organism with many integrated functions. As these important genes and sequences are identified, they can be modified, rearranged, or otherwise manipulated, and the effects on development can be analyzed. The study of unique kinds of mutants (called homeotic mutants) in the fruit fly, *Drosophila melanogaster,* have been invaluable in revealing how major developmental processes in this organism are controlled. Similar activity in plant developmental biology could produce comparably significant breakthroughs in the understanding of plant growth.

Plant cell culture is an additional tool with potential for facilitating advances in plant developmental biology. Some developmental processes can be controlled in cell cultures; a notable example of such a process is somatic embryogenesis, the development of embryos, and then plants, from cells in culture. This high degree of control and the capacity for manipulation that plant cell culture provides permit a developmental process to be intensively and effectively studied in an appropriate cell culture system. For example, mutant carrot cell cultures that are blocked at different stages of somatic embryogenesis are currently facilitating the greater understanding of this developmental process.

Further research in plant developmental biology will result in greater understanding of the crop plant as a biological organism. A number of specific areas can benefit directly from increased knowledge in this field. Yield is the sum of a number of components, many of which are based in plant form and function. The shape of a plant may critically influence its productivity. For example, new wheat and rice varieties with shorter stalks are already improving yields in many parts of the world. Branching patterns and leaf morphology can influence the severity of disease. Further understanding of the biological basis for plant form might permit more extensive manipulation of form with a resulting change in productivity. There is some information on the role that several of the recognized plant hormones play in communication between plant cells and between different parts of a plant. Additional understanding of the coordination of processes on an organismic level might also inspire the development of new plant growth regulators for the deliberate modification of agriculturally important processes such as senescence and assimilate partitioning (the distribution of photosynthetic assimilates to different parts of a plant). Finally, there is no question that the regeneration of plants from cell cultures of important crop species such as grain legumes and cereals will be improved by further knowledge and will thus facilitate genetic engineering with these crops.

PLANT AND PEST INTERACTIONS

Although dramatic successes in disease control and to a lesser extent in insect control have been achieved through a variety of plant-breeding techniques, the basic genetic, biochemical, and physiological mechanisms that enable plants to resist attack by pathogens and insects remain largely unknown. Effective chemical controls for diseases of plants caused by bacteria and viruses are lacking at present, and many chemicals used to control fungal, nematode, weed, and insect pests cause concern in terms of their safety to humans and the environment. In addition, pest populations have demonstrated the ability to acquire resistance to pesticides, pointing to the need for greater emphasis on the search for alternative approaches to minimize losses caused by plant pests.

The small size and relative simplicity of the genome in bacteria, viruses, and viroids—the latter being a very small pathogen, smaller than viruses—suggest that available molecular biology and genetic engineering techniques could be used to determine the factors that control virulence in the pathogen and thereby contribute to the development of more effective methods of control. The structure of the simplest pathogens, viruses, and viroids consists of no more than a nucleic acid sequence of DNA or RNA and, in the case of viruses, a protein coat. The key to recognition by a virus of a specific plant host and resistance to viral damage to the host probably will be revealed in the details of the role that the host's nucleic acid and protein-synthesizing systems have in replicating the virus. Viroids may be the first pathogen for which the disease-causing mechanism is known. Their RNAs have been sequenced and are small enough (a few hundred bases) so that specific point mutants can be constructed and the mechanism of their pathogenicity tested. One hypothesis is that viroids may interfere with the normal functioning of a host's small nuclear RNAs. These nuclear RNAs are thought to be involved in removing introns and splicing them out of host RNA transcripts.

Viruses are carried from plant to plant by vectors such as insects, mites, nematodes, or soil fungi. In a few cases, viruses multiply within insect vectors, as well as in the host plant. This remarkable versatility of replication contrasts with the specificity between host and virus illustrated by the failure of the same viruses to multiply in closely related plants or insects. The molecular basis for this highly specific recognition of divergent hosts is unknown, but viral mutants exist that have lost the ability to multiply in one of their hosts but not in others. Such mutants should be useful in investigating this problem.

Recent advances in bacterial genetics provide new tools for studying the underlying genetic and biochemical mechanisms that control virulence factors in bacteria. Knowledge of virulence factors will aid in determining mechanisms of plant resistance to pathogenic bacteria. A lack of understanding of the genetics of the attacking organisms has frustrated previous attempts to analyze the complexities of the biochemical reactions triggered when pathogens invade susceptible or resistant host plants. Most of the bacteria that cause diseases in plants are strains of *Erwinia* and *Pseudomonas.* A wealth of information is now available on the genetics of closely related bacteria such as *E. coli* and other *Pseudomonas* species, and exploitation of this information and use of new techniques are just beginning. Recent advances in studies with the crown gall bacterium *Agrobacterium tumefaciens* and its tumor-inducing (Ti) plasmid have elucidated the molecular biology of the interaction between this plant parasite and its plant hosts. Although crown gall is an atypical plant disease, results obtained using it indicate the possible advances that could be made with other pathogens.

The work with the Ti plasmid has progressed particularly rapidly because of its potential use as a vector for introducing foreign DNA into plant cells. Studies with other bacterial or viral pathogens also could lead to potential vectors for use in genetic engineering.

Certain groups of nonpathogenic bacteria have a remarkable ability to exist and multiply on leaf surfaces. Little is known about the role of such bacteria in aiding or antagonizing invasion of the host by pathogenic bacteria or fungi. It is possible that these bacteria can be used as antagonists of plant-pathogenic forms directly. Strains might be altered genetically to survive and antagonize plant-pathogenic bacteria and fungi.

Progress has been made recently in identifying the chemical structure of bacterial toxins and host-specific fungal toxins. A number of elicitors, compounds that trigger resistance reactions in plants, have also been identified. Key areas for research are those required to determine both the nature of receptors in plants that react to elicitor molecules, which are formed either by the pathogen or indirectly as a result of their invasion of the host, and the nature of the messenger molecules that are transmitted throughout the host plant and initiate the production of compounds involved in the resistance reaction. For example, one such compound, described as a proteolytic enzyme inhibitor-inducing factor, is released by damage to almost any part of a tomato plant; it spreads systematically throughout the plant and triggers the production of a compound that inhibits enzymes that digest proteins. It is likely that other compounds directly or indirectly inhibitory to bacteria and fungi can be formed in plants in a similar manner. Knowledge of such systems can provide clues to identifying resistance mechanisms in plants, can aid in the selection of resistant crop varieties, and can lead to the design of chemicals that elicit the resistance reaction.

In the case of insects, the relationship between host and herbivore is based on a behavioral response on the part of the insect to secondary plant metabolites. To understand these interactions it will be necessary to know the identity of the plant metabolites that stimulate or inhibit feeding by the insect, the characteristics of the insect's chemical sensory system that permit detection of different plant metabolites, and the nature and mechanism of any plant responses to insect attacks. For example, insect neurobiology studies are identifying new neurotransmitters, including some that are apparently unique to insects and may lead to new insect control methods. Also, many insects do the greatest damage while they are feeding at a larval stage. They are maintained in this stage by juvenile hormones. Some plants can produce compounds that interfere with the action of the insect juvenile hormone and result in an early termination of the larval stage, and thus reduce damage to the plant.

New techniques are now available for understanding chemosensory systems (e.g., electrophysiology, biochemistry of receptor proteins, and application of information theory to sensory coding and central nervous system integration) and for identifying plant volatiles, phenolics, alkaloids, and so forth (e.g., gas chromatography and mass spectroscopy). Organized collaborative efforts of plant product chemists, plant physiologists, and electrophysiologists should be aimed at identifying those plant chemicals that attract insects and induce feeding or oviposition and those compounds that prevent these activities by acting as deterrents and as feeding suppressants. Such approaches now hold great promise for the development of improved methods of control and for the selection of resistant plant varieties.

ORGANISMS IN THE RHIZOSPHERE

Soils are the primary source of nutrients for plants, animals, and people. Extensive research has been done on the physical and chemical nature of soils, but the microbial ecology of the rhizosphere, the soil-root zone, has been neglected even though some of the most destructive diseases of plants are those caused by soil-inhabiting organisms. These diseases are among the most difficult of all to control. Treatment with chemicals is costly, and effective compounds can often cause dramatic shifts in populations of nontarget organisms and may result in shifts to different pathogens, which would be similar to the unanticipated and undesirable shifts between beneficial and harmful insect populations experienced with the use of insecticides.

Major advances in the control of soil-borne pathogens have resulted from the development of resistant cultivars; however, additional progress in reducing crop losses from these pathogens may depend on the use of biological control systems, because effective resistance systems are not available. Some recent studies in the states of California and Washington on disease-suppressive soils indicate that certain strains of the bacterium *Pseudomonas* have the ability to colonize the surfaces of roots of a number of crop species, resulting in enhancement of root development and increased yields. Applications of such strains of bacteria to wheat seed have resulted in impressive yield increases on soils infested with the destructive fungal root pathogen that causes the "take-all" disease. It is possible that genetic engineering can be used to enhance the effectiveness of root-colonizing organisms as biological control agents.

Another major area for additional study involves the utilization of symbiotic fungi that form intimate associations with roots and enhance nutrient uptake and root development. These associations are called mycorrhizae. Two basic types of mycorrhizal fungus exist: the ectomycorrhizal forms that live on the outer tissues of roots and the endomycorrhizae that invade and live primarily inside root tissues. Ectomycorrhizae have been shown to be particularly important to forest trees, especially for seedling growth. The potential for large-scale exploitation of mycorrhizae through inoculation of tree seedlings with improved mycorrhizal fungal strains requires additional research on the biology, genetics, and ecology of soil organisms, as well as the study of which trees are physiologically best adapted to mycorrhizal relationships.

Endomycorrhizae are found primarily in the roots of annual and perennial grasses and shrubs, and their value to plant growth has been clearly shown. It has not yet been learned how endomycorrhizal fungi can be cultured and used as inoculants for seeds or seedlings, but advances in the knowledge of how to manipulate these ubiquitous and important fungi have great potential for beneficial effects on crop production.

Minimum- and no-till cropping practices have been used to counteract severe soil erosion conditions and to save on the fuel costs of operating farm machinery, but they leave organic debris on the soil surface, resulting in soil that is wetter and colder. Both factors have important effects on the population dynamics of soil organisms. In addition, the incorporation of fertilizers and organic matter into soil and the aeration and compaction of soil by tillage affect growth conditions for soil microorganisms in the rhizosphere and, in turn, crop productivity and efficiency. For example, as some microorganisms degrade surface debris they produce toxins that may inhibit seedling growth in the crop that follows. Also, evidence suggests that intensive tillage and heavy fertilizer applications have helped counteract suppression of crop growth caused by a variety of universal, debilitating but nonfatal root diseases. If beneficial bacteria such as the *Pseudomonas* strains noted above are to be utilized to reduce the expense of mechanical tillage and fertilizers, a better understanding of the ecology of the rhizosphere and interactions among soil organisms is needed.

The evidence of potential, beneficial effects from symbiotic mycorrhizal fungi and certain other beneficial root-colonizing bacteria and fungi highlights the need for a better understanding of the ecology and population dynamics of soil microorganisms and the effects of different cropping practices on these microorganisms. Some important areas of research include the role of soil microorganisms in nutrient cycling as they affect the availability of mineral nutrients to crops; elucidation of the inherent factors that determine the relationship of bacteria, fungi, and nematodes with plant roots; the genetics of soil microorganisms and the extent of genetic exchange among them; and the biological factors associated with those unique conditions that result in so-called disease-suppressive soils—soils in which pathogens are present but in which little or no crop root damage occurs. Techniques are now available for analyzing the biological and physical factors that would enable scientists to understand the basis for suppression of these pathogens.

Proper recognition of the ecology of soil microorganisms may lead to production of crops with less fertilizer, less soil erosion, and less tillage. Proper understanding of the ecology of the various organisms can result in decreased pest and soil erosion problems, which would permit development of farming systems with increased efficiency in crop production. Such systems might require more crop rotation, multiple cropping, minimum tillage, special green-fertilizer crops, new crops, or varieties bred for a specific cropping system. Basic information is essential for construction of the conceptual models required for rational manipulation of rhizosphere populations to minimize adverse effects of disease agents and to enhance the benefits from symbiotic organisms.

PHOTOSYNTHETIC ENERGY CONVERSION AND CARBON METABOLISM

Photosynthesis is the process on which all agricultural productivity depends. It is unlikely that significant ways for improving photosynthetic efficiency can be expected until the component steps of photosynthesis and their regulation in plants are understood. Understanding of the biological process could also be expected to help improve the efficiency of physical and chemical systems for solar energy conversion.

The first step in photosynthesis is the physical process in which light energy drives the separation of electrical charges. In green plants the charge separation occurs very efficiently, and the subsequent stabilization of the products of this reaction produces very high efficiency in energy conversion. With the availability of spectroscopic techniques to follow the movement of electrons at time resolutions in the subpicosecond range, the biophysical deciphering of the energy conversion process is well under way.

The stabilized photochemical products must be processed into chemical agents that can be utilized in biochemical processes within living cells. This processing occurs in solid-state devices—the photosynthetic membranes assembled within the chloroplast. There has been good progress in taking these membranes apart, purifying individual protein components, and recognizing their three-dimensional structure. So far only a few of the approximately 100 components have been studied completely, but the problem is manageable. The techniques of biochemistry, molecular genetics, and x-ray crystallography are equal to the task of taking the chloroplast apart, sorting, identifying, and understanding the pieces in terms of their individual shapes and functions, and then reassembling the pieces to validate the understanding of the whole.

Specific examples from the research accomplishments of the last few years demonstrate the potential for solving the structure of the photosynthetic apparatus. The enzyme (a carboxylase) that first captures carbon dioxide from the atmosphere to begin its conversion to carbohydrate is an exceedingly large and hard-to-handle aggregate of protein subunits. The smaller protein subunit piece is amenable to an analysis of its amino acid sequence. This has been accomplished by clever but conventional chemical techniques. The larger subunit piece of this enzyme is, by nature of its size, a forbidding task for chemical sequencing. The gene that directs the synthesis of the large subunit has been isolated, and this DNA was inserted into a bacterium so that it could be copied to an amount and condition that made further analysis much simpler. The sequence of nucleotide subunits that made up the genetic instructions that specify the sequence of amino acids in the large protein subunit was determined by an easily used enzymatic technique. By applying previously developed code translation techniques, the DNA sequence of nucleotides then could be translated into an amino acid sequence. This application of genetic engineering techniques simplified—and consequently shortened by many years—the task of determining the sequence of amino acids in the large subunit protein and the subsequent mapping of its enzymatically active sites; furthermore, these techniques can be applied in principle to each of the proteins involved in the photosynthetic process.

The next step needed is to learn the arrangement of this chain of amino acids in the subunits in three-dimensional space. An x-ray crystallographic analysis of the purified protein is under way and will tell the exact details of how the chain is folded and how the subunits are organized in a functional aggregate. Studies now in progress describe how the enzyme works and which parts of it are devoted to catalysis or regulatory modulation. While these studies are currently using chemical probing and alteration of the protein structure, it is clear that genetic manipulation (engineering) will make this process go much more quickly.

Finally, it is necessary to find out where this enzyme sits on the photosynthetic membrane and how it interacts with neighboring proteins that are its reaction partners. Because of its large size, the carboxylase enzyme can be recognized on the surface of the chloroplast membrane by electron microscopy. Certain techniques can be applied to identify neighboring molecules. When this work is complete, this component will be known just as the mainspring in a watch is known. The patient application of these techniques to the other parts of the mechanism is needed.

There is good evidence that many details of the basic process of photosynthesis are identical in all plant species. When the problem has been solved for one, it will have been largely solved for all. Furthermore, the energy processing that occurs in chloroplast membranes is very similar to energy processing in all types of cells. The detailed topographical analysis of the chloroplast membrane should provide powerful insights into energy conversion processes in every living system.

A wasteful reaction associated with photosynthetic reduction of carbon dioxide (CO_2) occurs in many plants. This reaction is known as photorespiration and can account for as much as a 50 percent loss in net photosynthesis. Recent research advances in photorespiration include the genetic verification of the biochemical pathway related to this process. Support of research on photorespiration in recent years has resulted in identifying some of the molecules that might be altered to eliminate or reduce photorespiration losses. Evidence points to the fact that the carboxylase enzyme that converts CO_2 to reduced carbon also carries out photorespiration when oxygen is present.

Research is needed so that photosynthesis can be understood beyond the chloroplast compartment where energy conversion and processing occur. Carbohydrate is the finished product of the chloroplast, and it must be exported to all parts of the plant. It is evident that there are intricate signals that regulate this export process. The filling up of storage reservoirs such as roots, tubers, or seeds with carbohydrates or the deposition of starch in the chloroplast can diminish the rate of leaf photosynthesis. Membrane transport regulators and enzyme modulators are being recognized as some of the signal mechanisms involved. It is likely that these processes can be elucidated and with considerable practical results, such as better controlling the overall efficiency of photosynthesis and guiding the distribution of products into the harvestable parts on the plant.

Understanding the partitioning throughout the plant of the carbohydrate products of photosynthesis should enable a directed approach to increase their movement into the part of the plant harvested. The importance of such an approach is supported by the fact that most of the yield increase in wheat over the last several decades is directly correlated with empirical improvement of assimilate partitioning into the seed.

Assimilate-partitioning research is ready for an expanded effort. Assimilates include sucrose, the major transportable product of photosynthesis, but also include other compounds such as amino acids and other nitrogen-containing molecules. Initial physiological and molecular understanding of the route of the transfer of sucrose includes proton-coupled sucrose cotransport driven by an ATPase to deliver the sucrose to the plant's circulatory system. Similar understanding of the energy-driven transfer in transporting sucrose into the seed has been developed. The compounds of nitrogen that are transported from leaves to the seeds are being explored.

A large, emerging problem area in photosynthesis has to do with understanding the energy costs of biological processes and the relation of these costs to the energy source in the chloroplast. Nitrogen fixation, efficient mineral uptake, and stress resistance all have an energy cost. The relation of benefits to costs in these processes must be optimized. This can only be based on an understanding of the chemical and physiological details of how the intact plants manage these interactions. The methodology for this understanding is available or nearly so, and thus it is simply a matter of applying more effort to solve the problems.

MINERAL UPTAKE, DEFICIENCIES, AND EXCESSES

As a crop plant interacts with its soil environment, a number of mineral stresses can impair its functions. While these stresses are clearly recognized, they are poorly understood. However, the tools are available for in-depth examination of these problems. Single-gene mutants that vary in response to mineral stress are known and could facilitate investigations into the fundamental mechanisms underlying mineral-related problems and their potential solutions. Increased research in this area would provide the knowledge needed to improve crop production and efficiency in affected environments.

A mineral stress can be imposed by too little of an essential nutrient, too much of a toxic element, or the imbalanced distribution of a nutrient within a plant. The major essential nutrients, nitrogen, phosphorus, and potassium, must often be applied to agricultural soils at great expense in terms of both money and resources. Additionally, some soils are also deficient in micronutrients (zinc or boron, for example). The efficient utilization of a nutrient is a function of the uptake, distribution, and ultimate utilization of the molecule. Current understanding of these processes is sketchy at best. That there is genetic variability in nutrient uptake and utilization is recognized, but not well understood. While a number of instances of single-gene control of differential efficiency of nutrient utilization are documented (for example, iron utilization efficiency in corn and tomato), neither the basis for the differential efficiency nor the nature of the gene or genes that control this efficiency is known. Mycorrhizal associations are significant factors in mineral nutrition, but much remains to be learned about the nature and regulation of these associations. (See section on Organisms in the Rhizosphere for discussion of mycorrhizae.)

The benefits to be derived from a greater understanding of nutrient utilization are twofold, namely, the more efficient production of crops on soils currently supplemented with high nutrient inputs (particularly nitrogen) via a reduction in inputs, and the extension of cropping to marginal soils deficient in essential nutrients. These improvements might be gained either through the development of genotypes that more efficiently utilize low levels of nutrients or perhaps by improved management practices that might result from a better understanding of the nutritional aspects of the agricultural ecosystem.

Mineral toxicities occur worldwide. In semiarid regions, salinity is a serious limiting factor. In the tropics, aluminum is the most serious mineral toxicity. Salinity affects perhaps 30 percent of the world's arable soils, and aluminum toxicity affects approximately 40 percent. In the United States, salinity affects many soils in the Southwest, and aluminum toxicity occurs primarily in the Southeast.

Similar knowledge is needed with regard to both of these major types of soil toxicities. For the most part, the fundamental mechanisms by which these toxic influences affect plant processes are little understood. As with nutrient efficiency, genetic differences in tolerance to these toxicities are well known but poorly understood. A number of mechanisms for tolerance are postulated. These include, for example, rhizosphere modification, ion exclusion, chelation, and compartmentalization. However, extensive investigation is needed to identify key mechanisms for particular crop-ion combinations. The energy cost of tolerance mechanisms is another consideration that bears investigation.

Among the important mineral toxicities, salinity has been a major subject of increased research activity. Recent genetic and physiological studies have yielded valuable insights. Comparable investigations into aluminum toxicity (and associated stresses in tropical soils) will also be productive. However, much more extensive study is needed in both areas before the nature of these stresses can be clearly understood.

The applications of knowledge about mineral toxicities are similar to those for mineral efficiency. This knowledge could be used to develop new crop genotypes better able to tolerate these toxicities. Such genotypes would permit efficient crop production on affected soils now under cultivation by reducing high-cost inputs (e.g., water leaching for salinity, liming for aluminum toxicity). Additionally, such genotypes could also be used to bring marginal land into production, thus extending the range of arable land. Increased knowledge might also result in improved cultural practices that might be introduced in areas affected by these stresses.

There is evidence that some disorders in crop plants involve deficiencies of nutrients—such as calcium—in specific organs, even though the overall concentration of the particular nutrient in the plant may be adequate. Recent findings indicate that calcium nutrition interacts with a number of other important influences on a crop (e.g., a large number of physiological disorders damaging to many fruits, root crops, and leafy vegetables during storage are related to their calcium content, and calcium also influences tolerance to salinity and aluminum toxicity). These findings underscore the importance of further study in this area.

BIOLOGICAL NITROGEN FIXATION

Research productivity in biological nitrogen fixation continues at a dynamic pace in chemistry, biochemistry, genetics, biology, and agronomy, with every indication that this pace will continue. Examples of recent research advances include sequencing of the nitrogen-fixation structural genes of several organisms, identification of regulating genes of nitrogen fixation, establishment of the similarity of the promoterlike sequence of algal nitrogen-fixation genes to plant chloroplasts, characterization of the molybdenum-iron cofactors of nitrogenase, identification of a nitrogen-fixing microsymbiont with nonlegume trees and the use of this microsymbiont *Frankia* to inoculate trees as the symbiotic bacterium *Rhizobium* is used to inoculate many legumes, increased yield of crops in specific geographic areas following inoculation with the bacterium *Azospirillum,* the enhancement of nutrient-ion uptake by *Azospirillum* suggesting opportunities to improve scavenging of limiting nutrients by crops, and the breeding of alfalfa-*Rhizobium* combinations with increased nitrogen fixation.

Nitrogen is a key limiting nutrient for crop productivity, with biological nitrogen fixation currently supplemented by 50 million metric tons of fertilizer nitrogen annually. Based on current technologies, it is estimated that the fertilizer nitrogen supplement will need to grow to 160 million metric tons by the year 2000. The fossil energy required to operate and the capital needed to construct fertilizer plants will be major expenses for agriculture. Applications from the expanding base of knowledge in biological nitrogen fixation may reduce the projected expansion in fertilizer nitrogen. Examples of highly probable short-term impacts are more effective *Rhizobium* inoculation technologies; legumes that are yield-responsive to fertilizers, thus increasing the harvest index for nitrogen; improved energy-efficient *Rhizobium* with genes to recapture some of the hydrogen that occurs as a by-product of nitrogen fixation; and *Azospirillum* inoculation of cereals to increase yields by presumably improved nutrient scavenging. Examples of less probable, midterm-impact technologies include *Rhizobium* that nodulates with insensitivity to the presence of high levels of fixed nitrogen and promiscuous *Rhizobium* to simplify inoculation. The more speculative, long-term impact technologies include achieving a major increase in photosynthate to fuel the symbiotic nitrogen-fixation system, finding catalysts that will produce synthetic nitrogen fertilizer with a zero direct energy input process, creating effective non-*Rhizobium* nitrogen-fixing associations, extending *Rhizobium*-legume-type symbiosis to nonlegumes, and transferring functioning nitrogen-fixation genes directly into crop plants. The last is one of the most challenging objectives of genetic engineering. The magnitude of the need for fixed nitrogen justifies the search for multiple alternate technologies.

PLANT GENETIC ENGINEERING

Initially, the most important contributions of the application of genetic engineering techniques to crop improvement may come not from splicing and recombining specific traits into crop plants, but rather from the enormous potential these new tools offer for acquiring a better understanding of the basic biology of plants.

The engineering aspects of genetic engineering consist of the biotechnical innovations that permit scientists to isolate genes from cells in the form of DNA molecules and to splice them into the DNA of other cells. As genetic information in the form of DNA is common to all living systems, the kinds of techniques developed and demonstrated with the DNA of bacteria and viruses also are applicable to the DNA of plants and animals. In effect, genetic engineering provides a new technique for recombining genetic information in plants by permitting the recombination to be done at the molecular level. This new biotechnology of genetic engineering is a powerful adjunct to plant breeding, which hitherto has been restricted by the limitations associated with manipulating whole organisms.

To identify and isolate a gene in the form of a DNA sequence, to clone copies of this DNA that can be inserted into a living cell, and to provide for the expression of this foreign DNA involve an enormous scope of effort. Exciting advances, all successfully demonstrated, include the isolation of plant genes, their characterization by DNA sequencing, their duplication through cloning, and their insertion as replicating foreign DNA into other plant cells. It is probably only a matter of a short time before the expression of a foreign, recombinant, plant-DNA gene is definitely demonstrated.

In spite of the significance of the foregoing developments, the genetic engineering method for transferring genes will not easily replace genetic recombination that can also be accomplished through conventional plant breeding. What, then, is a realistic assessment of the value of genetic engineering to agricultural research and to the improvement of crops?

To begin with, molecular biology and genetic engineering techniques have already provided unexpected details of and new insights into the organization and function of genetic traits. The direct demonstration of relationships between regulatory and structural genes, the presence of repeated sequences of DNA that occur with high frequency throughout the genome, the existence of noncoding sequences (introns) that must be processed out of the RNA transcript, and the surprising plasticity of the genome that was earlier thought to be rigidly conservative have all helped and forced scientists to think in new ways about the factors that influence and control the reassortment and expression of hereditary traits. Much of this increased understanding could only be derived through the use and analysis of the technologies associated with genetic engineering.

In terms of the direct application of genetic engineering to crop improvement, it is not possible to predict what specific practical advantages genetic engineering techniques will have over conventional breeding techniques in the near future. The application of this new technology to plants is still very much in its infancy, and there are many uncertainties about how crop species will respond to manipulation using the new biotechnology. It seems reasonable, however, to expect that there are traits controlled by a single gene or small groups of genes that can be transferred successfully using genetic engineering. Such transfers, which would otherwise not be possible through conventional breeding, might involve the exchange of DNA between noninterbreeding populations of plants or perhaps even between plants and viruses or bacteria. The end result would be to enlarge the genetic diversity available for crop improvement. This appears most probable for single-gene traits that convey resistance to herbicides and to some diseases, or tolerance to adverse soil conditions such as mineral deficiencies or excesses. The genetic engineering transfer of simple disease-resistance traits, however, seems impractical considering the rapidity with which pathogens regain virulence against single-gene resistance.

A major uncertainty in genetic engineering applications to plant improvement programs lies in the fact that yield and many of the desirable attributes of cultivated plants are conditioned by multigene systems in which the genes interact additively and epistatically. Epistatic genes alter the phenotypic expression of other, nonallelic genes. At present, the functional control of multigene systems is not understood. The optimization of such complex multigene systems for their most effective functioning may be beyond the current techniques of genetic engineering. Knowledge of the function of these multigene systems, however, is likely to be greatly benefited by molecular studies of gene expression.

Some groups of coordinately regulated genes are known to be organized in multigene families such as those controlling the synthesis of the storage protein in some seeds. While it represents an enormous task, some understanding of the biochemical functions encoded by such genes seems a prerequisite to the ability to enhance the characteristics that are determined by these multigenic systems.

Novel genetic variability can be generated by means of several plant cell culture techniques. Mutant selection in plant cell cultures has already yielded a number of new genotypes of interest, particularly in the area of herbicide resistance. Such mutants may have utility not only in agricultural production, but also as highly selectable markers for other genetic and molecular studies. Somaclonal variation (the *unselected* genotypic variation sometimes observed in plants regenerated from cultured cells) has already provided valuable new germ plasm in the case of sugarcane, and promises to do the same in other crops. The fundamental basis for somaclonal variation is currently unknown, and further study of this phenomenon may well reveal much about the nature of the plant genome.

Both mutant selection and somaclonal variation are technically simpler processes than is gene transfer. The technology required is less complex, and the difficulties associated with introducing a gene into a foreign setting are avoided. For certain breeding objectives, these two processes may be the preferred means by which to obtain the desired variability.

The assessment of the potential contribution of new genetic technologies in relation to that of conventional breeding will vary from crop to crop. For several major agronomic crops, very sophisticated and powerful conventional breeding technologies exist. But some important corps are very difficult to manipulate by conventional genetic means simply because they are difficult or impossible to reproduce sexually (sugarcane, banana, bamboo), or because their life cycle involves a long delay in reproductive maturation, which is true of most woody perennial crops. In these species, simple improvements that might otherwise be unattainable, in practical terms, might be achieved through genetic engineering and plant cell culture techniques.

The successful application of genetic engineering will generate new and previously unavailable sources of desirable genetic variability that can be exploited through conventional plant-breeding techniques. It has the potential for providing genetic variability that otherwise might not be obtainable through induced mutations or exotic races of the same plant species. Thus, interactions should be encouraged between scientists working with conventional plant-breeding techniques and those engaged in genetic engineering research in order to facilitate the interchange of ideas and material. It should also be kept in mind that even with many crops whose yields have been markedly increased by plant breeding, additional yield improvements still appear to be possible through conventional plant breeding, and that innovative adjuncts to conventional plant-breeding methods are still being discovered—for example, the use of meiotic mutants in *Solanum* species (potato) that result in unreduced gametes, which facilitates crossing potato species with different ploidy levels.

And finally, the positive side of the unpredictability of research in various areas of genetic engineering needs to be emphasized. The unexpected is part of the very nature of any discovery. One hundred years ago no one could have predicted the configuration of American agriculture today because no one could have predicted the scientific developments that would help shape that configuration. New knowledge acquired through advances in genetic engineering, together with the factor of nonpredictability, could likewise lead to major new developments in modern agriculture in the years ahead as research pursues its sometimes anticipated, sometimes unforeseen course. The knowledge of the biology of plants that is to be gained by aggressively exploiting these new research tools is a valuable end in itself.

RESEARCH BRIEFING PANEL ON NEUROSCIENCE

Floyd E. Bloom *(Chairman), Arthur V. Davis Center for Behavioral Neurobiology, Salk Institute*

Jack D. Barchas, *Stanford University Medical Center*

David H. Cohen, *State University of New York, Stony Brook*

John E. Dowling, *Harvard University*

Glen R. Elliott, *McLean Hospital, Boston, Massachusetts*

William Hodos, *University of Maryland, College Park*

Bruce S. McEwen, *Rockefeller University*

Charles Nicholson, *New York University Medical Center*

Donald J. Reis, *Cornell University Medical School*

Richard J. Wurtman, *Massachusetts Institute of Technology*

Staff

Fredric Solomon, ***Director,*** *Division of Mental Health and Behavioral Medicine, Institute of Medicine*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON NEUROSCIENCE

For this briefing, neuroscience was defined by the Office of Science and Technology Policy as "that body of research directed toward understanding the molecular, cellular, and intercellular processes in the central nervous system (CNS) and the way in which those processes are integrated in CNS functional control systems, with emphasis on research relating CNS functions with behavior."

Neuroscience is a relatively new and rapidly growing investigative field. One index of this growth has been the almost linear increase in members of the Society for Neuroscience—from 250 in 1971 to nearly 8,000 today. Programs of education in the neurosciences at the undergraduate, graduate, and post-doctoral levels have increased by 200 to 300 percent over the past four years. This growth has provided a sizable pool of trained investigators ready to pursue the opportunities for discovery within neuroscience.

Neuroscientists share overlapping interests, concepts, hypotheses, and methods of analysis. The field transcends traditional disciplinary boundaries and now includes many workers who have chosen to move out of their original scientific niches and into studies for which an interdisciplinary approach bridges many communicative gaps and allows for more rapid technical and conceptual discoveries. Given the pace of discoveries in this field, with its cascading flow of information, neuroscience has been recognized as an exciting field of fundamental investigation. The ultimate goal of neuroscience investigation is to understand the human brain. The opportunity to investigate seriously the physical basis of certain forms of mental activity is, thus, beginning to break down many of the traditional barriers that have existed between the biological and behavioral sciences.

Neuroscience research has enormous implications for health. The brain itself is a major potential site of diseases, such as the major neurological and psychiatric disorders. Moreover, the brain can also be the origin of disease states expressed by the rest of the body by generating behaviors, such as smoking or obesity, that can produce or exacerbate lung cancer, diabetes, or heart disease. Disorders due to brain damage—e.g., sensory impairment, convulsive disorders, mental retardation, stroke, and dementia—now afflict more than 25 million people. Another 2.5 million people suffer from severe mental disorders such as schizophrenia or major depression. Furthermore, the United States Public Health Service has estimated that at least 50 percent of the mortality from the 10 leading causes of death in the United States result from aspects of individual behavior. The health care burdens of behaviorally related problems continue to accumulate, because many of involved diseases are chronic, rather than fatal. Conservative estimates of these costs run to tens of billions of dollars each year.

Because dysfunctions of the brain's normal regulatory controls are major sources of illness, support for neuroscience has spread across many of the federal health research agencies. Lead agencies in funding neuroscience research are the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA); other agencies providing important support for neuroscience include the National Institute on Aging, National Eye Institute, National Institute of Dental Research, National Aeronautic and Space Administration, National Science Foundation, and Department of Defense. Overall, the 1983 level of support for neuroscience research is estimated to be less than $500 million, a dollar amount that offers less buying power than existed in the federal budget for this field a decade ago. The current level of investment is small in relation to the potential of such research for reducing the burden of illness and the specific scientific opportunities that exist now for accelerating progress.

TECHNICAL ADVANCES IN THE NEUROSCIENCES

This report describes some major opportunities for solving important problems in neuroscience. We first review some of the more recent technical and conceptual advances that now make these problems amenable to productive research. Other newly emerging methods, on the verge of wide application, also hold great promise, but the present report is a state-of-the-art view of a field in which important ideas have consistently followed the development of new investigative technologies. These advances are described according to the four major domains of inquiry in neuroscience research: molecular, cellular, intercellular, and behavioral.

MOLECULAR RESEARCH

Modern methods now permit the isolation and identification of specific molecules used by nerve cells to maintain their individual function as well as molecules that enable them to communicate with other neurons. Powerful new techniques have become available for investigating the molecular composition of the nerve cell membrane. Methods ranging from highly specific antibodies to physical probes of membrane fluidity provide information on molecules that are specific to certain classes of brain cells. Recombinant DNA techniques have recently been used to characterize the gene sequences responsible for neuron-specific products, especially certain of the neuropeptides that appear to transmit signals between nerve cells (neurotransmitters). The same methods may soon reveal the structure of their receptor molecules and of other components unique to nerve cells. The design of better molecular probes to provide a more complete chemical description of the nervous system is under way.

New analytical chemical techniques offer marked improvement in detecting, characterizing, and measuring substances within the nervous system. Such methods as mass spectrometry, nuclear magnetic resonance, high performance liquid chromatography, and radioimmunoassays can be used alone or in combination to detect substances at one-tenth to one-hundredth the concentrations needed with earlier methods.

Available evidence suggests that only a fraction of the neurotransmitters used by the brain are now known. These generally were discovered by seeking in the mammalian brain substances that were first encountered in other tissues, including the gastrointestinal system, the endocrine system, and neurons of invertebrates. It now is clear that an unexpectedly rich variety of chemicals are used for communicating within our brains—repeated time and again from strain to strain and from organ to organ. Neurotransmitters are now being discovered at a rapid pace due to the perfection of classical methods for isolating biologically active substances from tissue extracts and by the more recent application of the methods of modern molecular genetics.

The molecular mechanisms underlying the electrical activity of nerve cells also is being examined more thoroughly with the newer biophysical methods. Until recently, virtually all such information was derived from the use of microelectrode probes. Although much was learned from such studies, there were severe limitations in their ability to detect gradual electrical changes in the cell that generate neuronal action potentials. Furthermore, those electrodes were especially difficult to use on small neurons, which are the most numerous cell class in the brain. Within the past few years, a new recording technique has been developed as a result of advances in solid state electronics. It promises to provide information on membrane properties thought virtually unobtainable a few years ago. This technique, commonly referred to as "patch clamping," enables an investigator to record from the tiniest of neurons and detect the opening and closing of a single ion channel as a small patch of membrane responds to neurotransmitters. The method also enables the investigator to manipulate the internal environment of a single neuron. As more aspects of this technique are perfected, the data already available indicate the beginning of a new magnitude of discovery.

CELLULAR AND INTERCELLULAR RESEARCH

Much of the current excitement in neuroscience comes from an increased ability to understand the biology of the cells of the brain and the molecular processes by which they communicate and regulate each other. New methods for tracing circuits in the brain and for determining the molecular mechanisms underlying the transfer of information between nerve cells have greatly enriched our understanding of how the brain receives and processes information. However, the task of describing the connections of the brain and the actions of these connections is far from complete. To be fully appreciated, the newly identified circuits also must be considered in terms of the transmitter substances by which they communicate and the molecular mechanisms that mediate these signals. These aspects of intercellular communication are now capable of being studied together.

Expanded understanding of a biochemically specifiable cellular neuroanatomy has not only enlarged our views of how neurons communicate but has also permitted the development of new approaches to assessing and treating diseases of the brain and behavior. With every new transmitter isolated and identified, research provides a means to make analogs that can simulate or antagonize their actions, as well as methods to measure their relevance to unexplained diseases of the nervous system. Such studies also are far from complete.

Much useful information on the cellular properties of the nervous system is being gained through the use of cell cultures and other *in vitro* preparations. These are uniquely suited for studying such important issues as the definition of chemical and environmental requirements for neuronal maintenance and neuronal growth.

Genetic mutants are being used increasingly in analyzing the functional role of specific neurochemicals or structures and in modeling certain disease states. Typically, mutants are selected on the basis of a behavioral or neurological deficit and then studied for their underlying chemical or structural defects. For example, mice showing major deficits in postural orientation and locomotion have been shown to lack the normal cellular organization of a major brain structure, the cerebellum. Mice lacking receptors for the male hormone testosterone are found to be deficient in the ability to display certain male-specific behavior and to lack certain male-characteristic neural structures, notably in the spinal cord nuclei that normally innervate the penis. The process of genetic hybridization is permitting the development of animal strains with specific neurological or behavioral properties, ranging from resistance to alcohol intoxication to the ability to acquire new reflex skills. Such strains increase the chances of determining the cellular mechanisms underlying genetically expressed traits.

BEHAVIORAL PROCESS RESEARCH

In recent years the available repertoire of techniques for recording the activity of the brain during behavior has expanded greatly. Methods now range from the study of the activity of single nerve cells in animals to the ability to describe regional metabolic activity in human beings. Applications of these powerful methods are opening a new understanding of the neural basis of certain behavior. It now is possible to correlate the activity of specific neurons with such well-controlled behavior as learned and unlearned movements of the hand and the eyes, and thus to gain a high-resolution view of how specified arrays of neurons process information in the initiation and control of patterns of muscular contraction. In experiments for which an animal is taught to make or alter some behavioral response, studies now reveal the critical value in this process of systems by which the brain selectively directs its attention to the events of the internal or external world. This type of study begins to explore the changes in neuronal activity that reflect learning and the storage of information.

New research tools include such techniques as recording the brain's electrical activity without having to insert electrodes directly into the brain (so-called "noninvasive recording"). The development of imaging techniques, such as Position Emission Tomography (PET Scan) and Nuclear Magnetic Resonance scannings and of other direct functional measurements, such as event-related electrical potentials, have opened new opportunities for approaching questions about the neural localization and basis of higher cognitive function and perception in human beings. As normal patterns of brain activity are characterized, the potential of these techniques for detecting and identifying abnormal function increases dramatically. Remarkable advances in this context are at their inception. The capacity to span investigations from single-cell correlates of behavior to correlations between populations of neurons and behavior offers promise for an understanding of how arrays of neurons are organized to produce the ultimate output of the nervous system—behavior—and for exploring how such arrays of cells can be modified by experience and perturbed by certain environmental influences and disease.

New concepts and techniques in the analysis of behavior have made essential contributions to efforts to study basic neurochemical and neurophysiological mechanisms. Increased precision in recording of behavior has vastly improved studies of its neural and hormonal controls. Technical advances also have contributed to the better analysis of behavior. For example, telemetry with battery-powered, portable transmitters has made possible the study of freely moving animals in the wild and their patterns of social behavior. Similarly, video recording, which enables later multivariable scoring by several different observers, has led to sophisticated analyses of behavior ranging from aggression to mother-infant interactions.

COMPUTERS IN NEUROSCIENCE

The advent of powerful, inexpensive microprocessors and increased access to large computers has opened new avenues of research at all levels of inquiry in neuroscience. For example, anatomical data gathered with a variety of microscopic techniques, including precise tracking of specific molecules in selected cells, can be assembled into data bases in a computer and then reconstructed into three-dimensional representations and displayed as color graphics. In other applications, large populations of suitably labeled cells can be identified and quantified with computer techniques, thus enabling comparisons among normal, experimental, and disease-modified brain regions. Computers also make feasible the collection and analysis of simultaneous recordings of electrical activity of many neurons. On-line microprocessors are particularly useful for condensing, displaying, and storing such data. This immediate analysis of data can guide the experimenter in the choice of stimulus-response tests during the experiment.

Recent advances in defining aspects of behavior, together with video recording, permit a quantitative approach to behavior. In this context, the computer can be used to administer and modify behavioral models and to aid in codifying and collating data. In the latter context, video images can be analyzed frame by frame using interactive computer-driven graphic techniques, thus permitting very large assemblies of data to be handled.

The rapidly expanding availability of small computers and terminals, together with inexpensive data communication capabilities, makes feasible the establishment of networks of frequently updated data bases in many areas of neuroscience that groups of investigators can share as common repositories of information.

TARGETS OF OPPORTUNITY IN NEUROSCIENCE

Interest in neuroscience arises from efforts of many scientists to explore certain basic questions about the brain, its growth and function, and disorders that alter its ability to function properly. The opportunities selected for attention in this report relate to specific aspects of the same questions that have aroused the curiosity of scientists of the brain for decades. What now opens these questions to serious investments of experimental effort is the emergence of adequate tools, an accumulated background of fundamental information, and the trained manpower to make these efforts fruitful. Eventual success of the effort depends on the maintenance of research programs in the fundamental subdisciplines of neuroscience from which the present exploratory opportunities will be launched. Five major opportunities, selected because of these recent developments, offer extraordinary promise for promoting an understanding of the structure and function of the behaving brain, and its relevance to major health problems.

HOW DOES THE BRAIN GROW AND MAINTAIN ITSELF?

Contemporary neuroscience is now in a position to consider how the brain assembles itself from its primitive neuronal elements, how the neurons differentiate with specific biophysical and biochemical properties, how they seek out and connect with other neurons to form the adult brain, and how the neurons maintain themselves and their connections. Current efforts seek to establish whether neurons keep an ability to form new connections in adult life, in response either to environmental signals or to injury, and whether these properties are lost as neurons age. It now seems possible to determine how the orderly process of growth and differentiation is regulated by the fetal environment, by maternal factors, and by the mother's health and environment. This research is moving to the point of determining how external factors, interacting with gene expression, result in abnormal patterns of development and how such abnormalities result in the congenital brain defects affecting individuals with mental retardation or eventual mental disease. The field is in an explosive growth phase, not only for neuroscience but for biology in general. Three representative topics may be singled out: growth, differentiation, and maintenance of neurons.

It was once believed that neurons formed, made connections to each other, and developed into the adult brain in an orderly and genetically preprogrammed way. There is now evidence that these processes depend on or can be modified by chemical agents secreted by other cells and sensed by the neurons. Some factors are essential for neuronal survival when connections are first being made. Other factors may direct gene sequences needed to produce specific neurotransmitters. Still others may stimulate and direct the growth of neuronal processes to specific target cells. At present only a few factors have been identified, purified, and synthesized. Their further identification is hampered by their existence in only very small amounts. The techniques of recombinant DNA technology should further promote the identification and synthesis of such compounds and permit analysis of how they may influence the normal growth and development of nerve cells.

Neural Regeneration and Transplantation

The fact that mature nerve cells do not undergo cell division makes the brain particularly vulnerable to injury. Repair in nerves outside the brain can be achieved by regrowth of the neuronal processes. For example, severed nerves in the hand can undergo full regeneration. It has recently been recognized that cells of the brain also have the capacity to regenerate connections; however, some unknown factors apparently interfere with such growth. It has been proposed that such processes of regrowth may underlie the partial recovery of function that follows brain injury, stroke, or spinal cord damage. Present knowledge suggests that other specific tissue factors may promote growth of neurons. Additionally, in some experimental animals, certain neurons lost by degeneration or by congenital absence can be replaced by transplantation of comparable neurons from other normal animals. A preliminary attempt has already been made in Sweden to apply this transplantation approach to treat a patient with Parkinson's disease. These opportunities rest upon the development of anatomical techniques for identification of chemically specified nerve cells, improved techniques for studying regulatory communicative processes in tissue slices, and the increased accessibility of the techniques of molecular genetics for identification of small regulation molecules. We now have the opportunity to acquire further information about how such transplants can replace functions of absent cells in the brain, and possibly provide eventual applications to treating human disease.

Aging

Aging relates to the brain in two ways. First, some populations of nerve cells seem especially susceptible to dying with age, leading to the clinical features of the aging brain, dementia, and loss of mobility. The bases for these cellular losses are unknown. A second, more speculative relationship is that loss of a small population of brain cells, possibly endocrine-related neurons of the hypothalamus, may be the biological trigger that results in the intrinsic aging of tissue throughout the body. Perhaps one of the most promising areas in neuroscience relates to dementia, a problem destined to increase in social importance as the proportion of our aged population increases. Recent studies indicate that a specific group of cellular connections between the forebrain and the cerebral cortex appears to be lost in individuals with senile dementia. The transmitter for this circuit is believed to be acetylcholine. When drugs that interfere with acetylcholine transmission are given to normal subjects, many of the memory defects characteristic of the demented individual are produced. These findings are of great importance, for they may relate a specific cell population with a known transmitter to a disease of important economic consequences. The neurochemical identity of these cells raises the possibility of identifying subjects at risk by analysis of body fluids and offers the prospect of designing drugs that could replace actions of the lost neurons. Yet to be discovered are the molecular properties of these cells that make them particularly vulnerable to degeneration, the role of genetics in predicting vulnerability, and the possible role of environmental agents in the degenerative process.

HOW DOES THE BRAIN ACQUIRE, STORE, AND USE INFORMATION?

An important aspect of the brain's role as the organ of behavior is its ability to acquire information about the external and internal environments, to store and evaluate this information, and to take specific action based on it. Recent advances in the field of sensory motor integration enhance our understanding of how the brain regulates the flow of information it receives. Advances have been made possible by studies of events in the brain cells of awake animals engaged in monitoring and vigilance tests; of the precision of eye movement in animals and humans; and of attention, sensory mechanisms, and perception in animals and humans.

Other studies appear to be pointing the way toward solving the mysteries of how and where information is stored in the brain. Progress is being made on several fronts. At the cellular level, membrane specialists and neurochemists have begun to unravel the nature of the modifications that occur with time after a stimulus is received. There have been studies of memory and its neural correlates in animals ranging from invertebrates to nonprimate mammals, whose nervous systems are far less complex than those of human beings. Observed changes suggest that experience modifies neuronal communication through specific metabolic molecular events that control the physical properties of the nerve membrane and the secretion of neurotransmitters. Using the new culture techniques, it has been possible to begin studying the storage capabilities of neuronal networks contained within functional, isolated slices of mammalian brain. Other studies have investigated learning, memory, storage, and retrieval in animals, in normal children and young adults, and in people who have various types of memory disorders, such as the normal aged and those with Alzheimer's disease.

Two other promising subdisciplines deal with the ways in which the brain uses information. One of these is neuroethology, which studies the neuronal basis of behavior related to an animal's or person's interaction with the environment and with other organisms, such as courtship, parental care, establishment and defense of territories, aggression, and social communication. Many of these behaviors are genetically preprogrammed, but they also can be affected by specific experiences. The other is the field of cognitive neuroscience, which investigates a wide range of intellectual processes in normal and neuropathological human populations. This approach has a range from the study of event-related electrical activity in the brain in awake humans to studies of language, decision making, and problem solving.

HOW DOES THE BRAIN MONITOR AND REGULATE THE BODY?

In addition to generating mental activity, thoughts, and feelings, the brain also regulates the levels of key chemicals in the bloodstream and thereby influences the health of every cell in the body, including its own neurons. The brain receives chemical information from specialized sensor cells whose outputs vary with the concentration of such compounds as oxygen, glucose, or certain hormones, or with such physical variables as blood pressure or body temperature. Special groups of neurons then "decide" whether this component is appropriate and, if not, activate nerve cells capable of restoring normal homeostasis. Such accommodation occurs in any of three output channels: 1) neuroendocrine transducers, which release hormones into the blood stream or actively inhibit their release; 2) autonomic neurons, which control such events as the beating of the heart, the resistance of the blood vessels, the release of energy-rich compounds from the liver and muscles, and the secretion of some peripheral hormones; and 3) behavior, which might call for drinking water to restore blood volume, putting on more clothing to raise body temperature, or eating food to raise blood glucose. Disturbances in these sensor-integrator-effector loops may lead to a number of common disorders, including hypertension, sudden cardiac death, obesity, and infertility. An understanding of which brain neurons are involved—of their locations, their neurotransmitters, connections, and their responsiveness to circulating hormone levels—is likely to facilitate the development of effective treatments for these diseases.

The brain is selectively permeable to many circulating compounds other than hormones and key metabolic constituents. Recent evidence suggests that this permeability can be altered, with adverse effects, under circumstances such as uncontrolled diabetes or severe kidney or liver disease. An understanding of the dynamic relationships between blood and brain levels of nutrients and metabolites should provide insights into such diseases and into brain function in general and suggests ways to take advantage of these relationships for the treatment of the brain dysfunction.

The brain also receives a great deal of information from the lungs, heart, blood vessels, and gastrointestinal tract through nerves. Although the existence of these nerves has been known for a long time, much remains to be learned about the central interconnections and the neurotransmitters involved in their communication. It is now clear that disturbances in their functional control by the brain can lead to major medical problems, including asthma, sudden infant death, hypertension, sudden cardiac death, and possibly peptic ulcer.

Another finding in recent years is that the body's immune system is not as independent of the brain as previously believed. Psychosocial events such as loss of a spouse or other severe stresses can result in a temporary depression of immune responses to infection or foreign agents in the body. There are suggestions that these effects are mediated through the brain, and it should now be possible to identify the hormones, neurotransmitters, and neural systems involved in modulating the immune system.

HOW DOES THE BRAIN EXPRESS RHYTHMS, DRIVES, AND EMOTIONS?

Certain behavioral functions of the brain, such as sleep, arousal, aggression, sexual drives, eating, drinking, and even drug-seeking behavior, are now subject to serious investigation. Progress in understanding some of these already is substantial. For example, at least in animals, neural pathways related to sexual behavior, aggression, and sleep have been mapped; some of the relevant neurotransmitters and sites of hormone action also have been identified. Recent discoveries of a number of previously unsuspected neurotransmitters raise questions about their relevance to these behaviors and states. The discovery of brain substances (endorphins) that have properties like those of opiate drugs has led to a broad new line of inquiry into addictive states.

Hormones secreted by the endocrine glands are regulated by the brain through signals from the pituitary. In turn the endocrine hormones feed back to the brain evidence of their activity, which permits the brain to keep the system within normal fixed limits of activity. However, these same hormones also can influence mood and motivate behaviors; such effects, still incompletely explained, may influence the response of the brain to drugs used to treat mental and neurological disorders. Sex-related differences in behavior and personality are now being traced to the hormonal actions to which the brain is sensitive early in development.

Even such everyday activities as eating a meal are now being subjected to more precise research by breaking down the behavior into its components. The sensation of hunger, the decision to eat based on sensory or social inputs, the choice of what to eat, and the decision of when to stop eating each may involve particular sets of brain neurons. Any of those neuronal sets could be responsible for appetite disorders, such as severe overeating or undereating.

Biological rhythms are another promising area of basic research that may be fundamental to understanding and classifying many types of inborn behavior. For example, the daily cycle of sleep and wakefulness normally is synchronized to the periodic external environment by an internal biological clock. This physiological timing system shares the properties of similar circadian rhythms in a wide array of life forms, from single-cell algae to the entire plant and animal kingdom. In recent years, there have been major advances in understanding the physiological and behavioral processes that affect the timing of the sleep-wake cycle and other biological rhythms. Such rhythms would be of interest in any case, but artificial lighting and rapid forms of travel have produced unprecedented changes in the temporal environment. The short- and long-term consequences of such shifts remain to be assessed adequately. The most important periodic environmental stimulus for entraining daily rhythms in nearly all species appears to be the alternation of environmental light and dark. A specific area of the brain has recently been identified as one pacemaker responsible for the generation of many of these rhythms.

An appealing feature of research on discrete, inborn brain states or drives is its potential application to more complex brain functions, such as memory and voluntary behavior. In relatively simple invertebrate animals, it now is possible to trace the entire neural circuit for such a behavior as orienting to a source of food. In theory, at least, investigators should be able to monitor the electrical activities of the cells in that circuit, to identify and study the neurotransmitters involved, and to study the effects on the system of specific environmental influences. Information gained from such studies should go a long way toward providing investigators with the tools needed to examine in detail more complex behaviors in higher animals, including human beings.

WHAT SPECIFIC ASPECTS OF BRAIN FUNCTION ARE ALTERED IN DISORDERS?

Rapid strides have been made in improving the diagnosis and treatment of some severe mental disorders, neurological conditions, and addictive states. The improving armamentarium of basic science has made it increasingly possible to examine these problems by means of human and animal models.

Biochemical Studies in Human Beings

Knowledge about recently discovered neurotransmitters will undoubtedly influence future advances in the understanding of psychiatric, neurological, and addictive states. Although further basic information is essential, the two dozen substances already regarded as important in communication between nerve cells make possible highly relevant studies of the role of specific neurotransmitters in various disease states. For example, investigators can assess the activity of various neurotransmitter systems by measuring substances in body fluids; this offers unprecedented opportunities for research into such disorders as depression and schizophrenia. Also, brain tissue taken after death is providing invaluable information about concentration and localization of neurotransmitters and related constituents, particularly in degenerative disorders of the CNS, such as multiple sclerosis.

Studies of the neurochemistry associated with neurological disorders and addictive states require not only substantial neurological and analytical chemical skills but also clinical research centers in which patients with various illnesses can be studied in controlled settings and with careful behavioral monitoring. Such facilities are rare, but their presence is essential.

Better techniques for studying brain functions also have clear implications for understanding other severe brain disorders. For example, Huntington's disease is a hereditary disorder characterized by involuntary movements and a deterioration of cognitive function. In most individuals with Huntington's disease, symptoms do not appear until well into their reproductive years, so that they must decide whether or not to have children without knowing if they will be transmitting this devastating disorder. Strides have been made in characterizing the anatomical and neurochemical changes that occur with this disease. Such information is hoped to help in developing early ways of detecting people who have Huntington's and in devising more effective treatments for its crippling effects. A similar strategy may be generalizable to other inheritable degenerative disorders of the nervous system that have more complex genetic and environmental contributions.

The new and still primitive noninvasive techniques for studying human brain function could lead to recognition of new categories of illness. Such methods are greatly expanding the potential for precisely localizing sites of brain damage and correlating such damage with changes in body function and in behavior. Available evidence already suggests that there are several forms of epilepsy, depression, and schizophrenia. In some cases clinically indistinguishable subtypes of disorders may prove to have quite different patterns of changes in the brain. There are suggestions that damaged areas in the temporal lobes of the cortex can appear as abnormal behavior, such as episodic rage or psychotic thinking, rather than as classical epileptic seizures. Identification of such subtypes will be critical in designing more effective treatments, in developing strategies for prevention, and in formulating and testing hypotheses about the underlying causes of disease.

Other Information Needed on Human Problems

Additional fundamental information is required about behavior and about more traditional biology. Improved rating skills that for such specific components of mental activity as memory, speech, mathematics, emotionality, stress, pathological fears, or thought processes are important examples. An ability to rate changes through time, including quite subtle changes, will enhance studies of the natural history of disorders in these abilities and of their treatment.

Behavioral studies that tie into basic biological work, such as investigation of sociophysiology—the ways in which basic social processes alter physiology—may provide information that will result in better controls and more careful clinical studies, as well as being directly germane to studies of the disorders themselves. The need for animal models of mental, neurological, and addictive disorders continues to be pressing. Although human studies are essential, there are no substitutes for animal studies in many types of investigations. Thus, the recent development of an animal model for the sleep disorder narcolepsy provides a new opportunity to study a vexing and difficult process.

EMERGING TARGETS OF OPPORTUNITY

Aside from the major opportunities for advancing knowledge within neuroscience, there are certain areas of promise for which the base of fundamental information and available methods is not yet sufficiently advanced to predict definitive results over the next few years. In almost all cases, these emerging areas of opportunity need a longer-term commitment because of the complexity of the problem area. In this category, we may mention the following.

1. What research strategies will clarify the currently unsolved neurological and mental disorders? Most modern concepts of CNS disease have rested on the hypothesis that some part of the brain's "hardware" has become pathologic—either a loss or perhaps a hyperactivity of a specific class of neuron. Recent advances in the ability to separate and purify the receptors of neurotransmitters and to follow their changes in the presence of chronic drug treatments suggests that there may be whole classes of "software" diseases which derive from abnormal patterns of receptor regulation. Also, greater insight into the nature of these receptors may well lead to the development of drugs that have been molecularly tailored to correct specific defects in cellular communication.

2. How does the brain deal with environmental toxins and drugs? Much has been learned about the ways in which brain cells respond to demands on their activity, but less is known about the ways in which toxic materials of various sorts produce their effects on the body, effects manifested in part or in whole through the brain. Especially critical are the effects produced by such substances as narcotics, alcohol, excess food leading to obesity, and cigarettes. These potential voluntary toxins account for a very large health burden, particularly costly because of the chronic nature of their abuse and the long lags between the adoption of the voluntary patterns of abuse and the appearance of adverse consequences to health. This includes possible effects on the developing fetal nervous system that may not be manifest by neurological or behavioral abnormalities until well into postnatal life.

3. How does the brain deal with infectious agents? Recent research has demonstrated that many degenerative disorders of the nervous system can be simulated in experimental animals by inducing immune responses against one or another neural component. These studies suggest that such responses could be triggered by the changes accompanying infection. Severe, lethal degenerative disorders of the brain are known to result from virus infections which occurred long ago, in which the virus has lain dormant in the nerve cells awaiting proper conditions for replication and tissue destruction. Some of the sites by which viruses such as the rabies virus can invade nerve cells may depend on the specific arrays of neurotransmitter receptors that a

susceptible cell carries. Some recent data suggest a viral basis for some forms of age-related dementia and for certain rare types of schizophrenia. Other diseases for which virus-related clues have been raised, for example, multiple sclerosis, are not yet to this endpoint. Nevertheless, determination of the way in which the brain typically is protected from but can succumb to infectious agents is an area of immense importance for additional future investigation in which advances at the fundamental levels already in progress will be likely to pave the way.

SUMMARY

Neuroscience is a relatively new and rapidly growing investigative field that transcends the traditional boundaries of fundamental scientific disciplines and encompasses research efforts in all clinical disciplines relating to diseases of the brain and behavior. This briefing report has described some of the basic technical and conceptual advances that have fueled the recent explosive advances in neuroscience research. Given the availability of a critical pool of skilled investigators—steadily supplemented by new investigators in training and by the influx of experienced investigators entering the neuroscience field from other disciplines—important opportunities now exist for significant enhancement of research efforts. The Panel has selected five broad targets of opportunity that are likely to experience advances in the near-term future. These targets of opportunity are crystallized by the following five general programmatic questions:

1. How does the brain grow and maintain itself?
2. How does the brain acquire, store, and use information?
3. How does the brain monitor and regulate the body?
4. How does the brain express rhythms, drives, and emotions?
5. What specific aspects of brain function are altered in disorders?

The Panel believes that, with additional resources, the techniques and skills now available could be put to immediate and productive use on these broad concerns. The product of such increased efforts not only would yield important intellectual advances in man's ability to understand the function of the human brain but also would offer direct insights into the important unsolved neurological and behavioral disorders, which by their early onset and chronic course, represent an extremely costly burden of illness to the nation.

The skills and tools we have highlighted are far from being universally shared. The Panel has not attempted to address the practical realities of making available the recently developed, often expensive investigative tools described, nor have we considered how scientists not yet engaged in neuroscience may be able to acquire the necessary research skills to enter the field. Although scientists in the field have expressed no consensus as to which areas may experience the next major advance, the five targets of opportunity to which this report calls attention represent areas of research to which the creative curiosity and individual efforts of many neuroscientists can be applied.

This report has emphasized areas that lie at the frontiers of neuroscience. It must not be forgotten that our present concepts of the structure and function of the brain have been built upon decades of painstaking, detailed research, during which time United States neuroscientists have been the leaders. Continuation of their fundamental efforts will be necessary if we are to avail ourselves of the opportunities that lie ahead.

RESEARCH BRIEFING PANEL ON HUMAN HEALTH EFFECTS OF HAZARDOUS CHEMICAL EXPOSURES

Norton Nelson ***(Chairman)****, New York University Medical Center*
Bruce N. Ames, *University of California, Berkeley*
Earl Benditt, *University of Washington*
Theodore Cairns, *Greenville, Delaware*
Ronald Estabrook, *University of Texas Medical School (Southwestern)*
Bernard Goldstein, *University of Medicine and Dentistry of New Jersey*
Joshua Lederberg, *Rockefeller University*
Brian MacMahon, *Harvard University School of Public Health*
Richard Merrill, *University of Virginia School of Law*
Vaun A. Newill, *Exxon Corporation*
Gerald Wogan, *Massachusetts Institute of Technology*

Staff

Robert G. Tardiff, ***Executive Director,*** *Board on Toxicology and Environmental Health Hazards, National Research Council*
Warren Muir, ***Consultant***
Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON HUMAN HEALTH EFFECTS OF HAZARDOUS CHEMICAL EXPOSURES

INTRODUCTION

This report is limited to research on the human health effects of hazardous exposures to chemicals, both man-made and natural. Although many physical factors, including ionizing and nonionizing radiation, are of public health importance, they are not included here. Nor did the Panel consider effects on the environment.

It should be recognized that personal habits, including cigarette smoking, unquestionably have major health impacts; and other factors, such as dietary constituents, may also be important. The initiatives proposed here should contribute a better assessment of the nature, variety, and extent of chemical threats, whatever their source, to the human population.

The extent of exposure is an indispensable component in any evaluation of the effects of hazardous chemicals. Research on means for estimating exposure by environmental monitoring, although very important, is beyond the purview of this Panel's study. One major proposal, however, deals with the estimation of individual exposure by the use of biochemical markers.

Many decisions made by the public and private sectors in relation to the control of risk of chemical exposures have had important economic consequences. Federal agencies and industry have often been compelled to make such decisions in the face of large scientific uncertainties. No amount of research can eliminate all uncertainties associated with assessing the risks of exposure to potentially hazardous chemicals or eliminate the controversial judgments inherent in any decision about control of chemical exposures. Yet there are opportunities to reduce scientific uncertainties and to increase the public's confidence that health is protected and that the economic consequences of imposed chemical control are justified. Recent dramatic advances, especially in molecular biology, have opened up the possibility of making substantial progress in reducing the scientific uncertainty in such decisions. Some of the best opportunities are of a fundamental nature that are likely to greatly enhance the ability to directly determine individual exposure, preclinical signs of disease, and chemically related illnesses in humans. Other promising efforts offer the prospect of greatly improved interpretation of laboratory data on possible chemical effects.

Of the many possible areas of potentially important and fruitful research, this Panel is recommending six that offer particular promise and are uniquely timely. Other important areas examined by the Panel included neurobehavioral effects, immunotoxicity, and organ-system studies, but they were eliminated from the Panel's primary recommendations at this time. These six areas were chosen based upon the Panel's belief that they offer the greatest prospect for making advances in our understanding of scientific issues that are likely to be pivotal in chemical risk decision making. Some of the initiatives relate to the improvement of existing assays; however, until such improvements are effected, existing methods must continue to be used. Some of the initiatives indicate that substantial progress can be made with research on several endpoints using human cells and tissues; however, the Panel recognized that several technical and legal considerations need to be resolved before such tools can be fully implemented and that for many other endpoints—beyond those covered in this report—substantial technical difficulties are anticipated before viable assays become generally available.

The six topical areas, for which priorities were not established, are:

I. Research to improve the basis for dose and interspecies extrapolations to humans.

II. Research into prediction of the effects of multiple chemical exposures.
III. Improvement in approaches to the determination of reproductive and developmental effects of chemicals.
IV. Development of cellular and molecular markers of exposure to chemicals and markers of preclinical effects.
V. Development and evaluation of means to distinguish among chemical cancer-causing agents on the basis of modes of action.
VI. Expanded use of existing federal data-collection activities for assessing human health effects.

I. RESEARCH TO IMPROVE THE BASIS FOR DOSE AND INTERSPECIES EXTRAPOLATIONS TO HUMANS

NEED: Control decisions for human health are based largely on extrapolation from tests on laboratory animals. The qualitative and quantitative biologic differences between test animals and humans vary substantially. Therefore, improvement in making such extrapolation is needed.

RECOMMENDATION FOR PROMISING RESEARCH: Comparative studies—e.g., on enzyme patterns and repair mechanisms—that use, in the case of humans, surgical and autopsy tissues.

ANTICIPATED IMPACT OF RESEARCH: Research to establish the qualitative and quantitative differences over a wide range of doses between test animals and humans can improve toxicologic predictions.

RATIONALE: The information available to make judgments about the possible effects of exposures to chemical substances often consists only of results of laboratory tests on species other than humans. Judgments made on the basis of such information are inevitably weak, because of a lack of understanding of the factors determining interspecies similarities and differences in responding to chemical agents.

Interspecies variations in responses to toxic chemicals can arise from differences in many characteristics, such as genetic inheritance; enzymes that activate and inactivate toxic materials; intracellular pathways of toxicity; membrane biochemistry and receptors; absorption, distribution, storage, and excretion; and physiologic differences in particular organs. Through the study of these differences between humans and test species (including the use of organs and cells in culture), it should be possible to improve greatly our ability to interpret laboratory data in terms of potential risks to humans.

Exacerbating the problem of dosage extrapolation and interspecies comparison is the increased heterogeneity of the genetic mix as one proceeds from purebred, genetically homogeneous strains of mice to the human population. Genetic composition and expression are dominant factors in determining the cellular response to chemical insult. Knowledge of regulation of expression is still rudimentary, but some methods to evaluate differences in "genetic potential" are now available. Such methods must be used, if we are to recognize and evaluate differences in susceptibility arising out of human genetic diversity.

Improvement in the scientific base for dosage and interspecies extrapolation underlies improvement in the basis for judgments about chemical-related health risks to the human population. Such improvement is now possible through exploitation of increased understanding of a number of areas, including gene expression, DNA repair mechanisms, enzyme induction and specificity, developments in cell-culture techniques, and advances in cell biology.

A number of specific subjects of particular promise that should be pursued are noted below:

1. Differences in bioactivating and protective enzyme systems and of repair systems should be studied to increase understanding of differences in pathologic response between humans and test species.
2. Human and laboratory animal genetic homogeneity should be explored to gain the knowledge needed to recognize and evaluate differences of susceptibility arising out of human genetic diversity and to enhance extrapolation of laboratory results to susceptible members of the genetically diverse human population.
3. The role of physiological characteristics, such as lung structure and function, should be pursued in concert with molecular and biochemical explorations.

Thus, it is recommended that a focused program of research be undertaken to improve the basis for dose and interspecies extrapolations of chemical risks to humans.

II. RESEARCH INTO PREDICTION OF THE EFFECTS OF MULTIPLE CHEMICAL EXPOSURES

NEED: Most chemical hazards involve concurrent or sequential exposure to multiple chemicals, e.g., product mixtures, toxic wastes, cigarette smoke, and pollutants in air, water, and the workplace. The scientific basis for evaluating health effects of such mixed exposures at often widely differing dose levels is extremely primitive.

RECOMMENDATION FOR PROMISING RESEARCH: Study of mechanisms by which chemicals alter biologic responses to yield addition, inhibition, or synergism of effects.

ANTICIPATED IMPACT OF RESEARCH: Such research can substantially improve ability to assess the health effects of environmental chemical exposures.

RATIONALE: Although study and control of environmental chemicals have necessarily focused on individual compounds, it is obvious that humans are exposed to a multiplicity of agents in their daily activities. Most chemical products in commerce and nearly all environmental pollutants—whether in the workplace, from hazardous-waste sites, water-effluent streams, or formulated products—are complex mixtures. To the extent that the effects of chemical mixtures differ from those of their separate components, there is an important need to develop approaches to the prediction of the effects of the mixtures, so that the basis for making control decisions about exposures to them can be improved. The needed studies can be conveniently divided into two types: 1) those focusing primarily on interactions among several potentially hazardous environmental chemicals, and 2) those evaluating interactions among environmental chemicals and other agents, such as foods and drugs.

Toxic chemicals can produce independent, additive, synergistic, or protective effects. In some situations, not only the extent but also the type of toxic effect of several chemicals together differs from the sum of each of the chemicals acting alone. Combined exposures can result in effects in different organ systems or in different toxic effects. Our understanding of the mechanism of action of individual agents has in a few instances begun to allow the prediction of effects of multiple agents, but much more information is needed.

Considerable data are now available on the effects of exposures of laboratory animals to specific chemical agents with regard to their abilities to activate or deactivate other chemicals. Studies in this area should be extended to the human as an aid to elucidating the underlying sources of the varied sensitivities of persons exposed to toxic chemicals.

Future studies should emphasize not only the biotransformations that lead to interactive effects, but also other processes, such as the absorption and distribution of chemicals. For example, with respect to the effect of one chemical on the absorption of another, competition for receptor sites in cellular membranes may be important in hormonal mediation. Such processes should be studied to determine the circumstances in which competition may be important.

An overall emphasis in future studies, wherever possible, should be on developing an understanding of general classes of compounds as an aid in predicting the nature of interactive effects. Such studies should complement the research on dosage and interspecies extrapolation described in Section I.

The results of all these studies would be valuable in guiding decisions about the many complex mixtures in commerce and in the environment. They would be extremely valuable in establishing priorities for the cleanup of multichemical hazardous-waste dumps.

Thus, it is recommended that there be a substantial increase in research into ways of predicting the effects of multiple chemical exposures at various dose levels.

III. IMPROVEMENT IN APPROACHES TO THE DETERMINATION OF REPRODUCTIVE AND DEVELOPMENTAL EFFECTS OF CHEMICALS

NEED: Current means for evaluating chemical injury to the reproductive function are of limited breadth and uncertain reliability, as are means for assessing chemical injury to the newborn.

RECOMMENDATION FOR PROMISING RESEARCH: Expanded research aimed at understanding mechanisms of chemical injury to the reproductive function.

ANTICIPATED IMPACT OF RESEARCH: Such research is likely to increase the reliability of experimental approaches and assessment techniques for evaluating chemicals that may cause reproductive abnormalities or birth defects.

RATIONALE: Various incidents of the past few years have increased public awareness of the potential for reproductive and developmental problems associated with exposure to hazardous chemicals—for example, the serious birth defects associated with thalidomide use during pregnancy, male sterility resulting from dibromochloropropane (DBCP) exposure in the workplace, vaginal cancer and adenosis in the daughters of women who were treated with diethylstilbestrol (DES) during pregnancy, fetal alcohol syndrome in the offspring of women consuming alcoholic beverages during pregnancy, and increased risk of spontaneous abortion in women who work in surgical operating rooms. As a result, both private and government organizations have been compelled to address potential reproductive and developmental effects of chemical substances to which humans are exposed. But the significance of the effects observed in current reproductive and developmental toxicity tests with respect to humans is unclear; accordingly, the control actions that should be instituted are uncertain.

The development of the ability to determine heritable changes brought about in the genomes of sperm and egg by exposures to toxic substances constitutes a pressing problem, because injury may persist and be expressed in later generations. The Panel believes that advances in molecular biology will help make possible the direct examination of regulation and expression of the human genome and its modification by chemicals.

The Panel has identified several areas of scientific inquiry that may contribute to the development of reliable tests for reproductive and developmental injury. These include research into:

1. Ways in which chemical agents lead to male and female infertility;
2. Morphologic and biologic characterization of sperm as a basis for detecting and understanding the action of environmental agents;
3. Physiology of implantation and the early stages of the maintenance of the fertilized ovum, particularly as related to potential vulnerability to chemical agents;
4. Prediction of the delayed effects of chemical agents on offspring;
5. Relation of postnatal adjustment, growth, and maturation to chemical exposure; and
6. Reproductive ability of groups of humans exposed to certain chemicals, especially at high concentrations.

With the expanded information that would be provided from research on these subjects, it should be possible to develop more specific and more interpretable tests, whose results could shape preventive measures to reduce the likelihood of reproductive and developmental effects of environmental chemicals.

Thus, it is recommended that federally supported research into the mechanisms of chemical induction of developmental and reproductive effects be increased so as to improve the assessment of the potential of environmental chemicals to cause these effects.

IV. DEVELOPMENT OF CELLULAR AND MOLECULAR MARKERS OF EXPOSURE TO CHEMICALS AND MARKERS OF PRECLINICAL EFFECTS

NEED: Sensitive means for estimating individual exposure to chemicals and for detecting early (preclinical) effects are urgently required.

RECOMMENDATION FOR PROMISING RESEARCH: Development and validation of techniques of high sensitivity and specificity for detecting the interaction of chemicals with, and their effects on, tissue components (DNAs, RNAs, hemoglobins, and other proteins) in humans for use in assessment of chemical exposure and risk.

ANTICIPATED IMPACT OF RESEARCH: Such procedures, if fully developed, would be powerful tools for direct evaluation of human exposure to environmental chemicals and its impact on human health.

RATIONALE: Current decision making about potentially hazardous chemicals is handicapped by the lack of anything more than fragmentary data on the levels of human exposure. Current approaches rely on projections derived from exposure models or on data from fixed monitoring stations. Generally, little information is available to measure or verify actual human exposure, especially to individuals.

Specific biologic markers of human exposure (as an adjunct to environmental monitoring) and of their preclinical effects offer a great opportunity to improve the ability to assess the effects of chemicals to which humans may be exposed.

There are several examples of promising research that should be pursued to develop and validate new clinical methods that would enable more accurate description of human exposures and detection of preclinical effects. Such methods should also be validated for animal systems so as to strengthen animal to human extrapolations. Promising methods that should be pursued include:

A. Means for evaluating individual exposure:

1. Analyses of expired air, blood, and urine that can be coupled with tests of toxic activity—for example, the application of short-term mutagenicity tests (e.g., the Salmonella/microsome test) to provide, inexpensively and rapidly, information on a portion of the body burden of mutagens.
2. Analyses for covalent and noncovalent binding of toxic chemicals to proteins or other constituents of assayable fluids or cells to be used as indicators of low-level exposures.
3. Highly sensitive and relatively rapid immunologic and chemical assays for identification of chemically modified nucleic acid bases that are excreted in the urine.
4. Use of monoclonal antibodies to detect specific chemical antigens for rapid screening of potentially exposed populations.

B. Means for detecting early (preclinical) effects:

1. Analyses of chromosomal alterations to detect genetic damage in peripheral blood cells (lymphocytes) of large exposed populations.
2. Detection of toxic effects on the reproductive systems from studies of changes in sperm morphology, including the use of special techniques, such as staining of DNA with fluorescent dyes and the use of rapid-flow cytofluorimetry to detect changes in the DNA organization of the sperm head.
3. Use of recent advances in gene cloning and DNA sequencing to study alterations in the genetic makeup of human cells directly.

With the further development and validation of these indicators, exposure assessments in humans will be strengthened; and these improved assessments will substantially increase the power of epidemiologic studies. Such measures will also improve the basis for extrapolation from laboratory exposures to human exposures. Potential applications include assessment of the consequences of human exposure in industrial and community settings and in connection with toxic-waste dumps.

Thus, it is recommended that support be increased for the development and validation of cellular and molecular markers of exposure to chemicals and of markers of preclinical effects of chemicals.

V. DEVELOPMENT AND EVALUATION OF MEANS TO DISTINGUISH AMONG CHEMICAL CANCER-CAUSING AGENTS ON THE BASIS OF MODES OF ACTION

NEED: Cancer induction in humans may often be the result of a complex series of chemical and biologic interactions. Yet, lacking full understanding of these processes, we continue to treat chemical causation of cancer as if it were a one-step process.

RECOMMENDATION FOR PROMISING RESEARCH: Research aimed at refining understanding of the stages of carcinogenesis, including "initiation" and "promotion," and the further exploration of possible fundamental differences, such as those between "genotoxic" and "nongenotoxic" carcinogens.

ANTICIPATED IMPACT OF RESEARCH: Such research may make it possible, in risk-assessment and -control decisions, to distinguish among potential cancer-causing chemicals on the basis of their modes of action.

RATIONALE: Current evidence indicates that carcinogenesis is a complex process that often involves the induction of a stable and permanent change in the genetic information of a cell (initiation), followed by successive stages that lead to the development of a malignant tumor. A number of factors, including chemical exposures and nutritional status, are known to affect the progression of initiated cells to tumors, either enhancing (promoting) or inhibiting tumor development. If the different modes of action of various chemicals in carcinogenesis were better defined and understood, it might substantially affect the way in which potentially hazardous chemical exposures are evaluated and acted upon.

Too little is known about the importance of man-made and natural environmental agents that may promote or inhibit tumor development. It is evident that man-made chemicals are not the only initiators or modulators of neoplastic changes. Some constituents of natural foodstuffs, uncooked or cooked, can act as initiators, promoters, or inhibitors of tumor formation; such naturally occurring factors may be important contributors and should be intensively explored.

A few chemicals—such as saccharin, TCDD, phenobarbital, PCBs, some chlorinated hydrocarbons, and bile salts—have been shown to induce tumors in animal tissues but show no evidence of genotoxicity; therefore, they may be acting through the promotion of cells previously initiated spontaneously or by unidentified agents. The number of environmental chemicals that may have similar promoting activity is unknown, because of the lack of bioassay systems for their detection. Recent advances in the fields of molecular and cellular biology offer prospects for obtaining information on the mechanisms by which promoting agents exert their effects. An understanding of the mechanisms of action of such chemicals is required for the development of appropriate test systems and the evaluation of their public health significance.

Substantially expanded research that may lead to improved understanding of the significance of promoting agents in human cancer incidence is recommended. Similarly, research should be performed to understand the mode of action of chemicals that may induce cancer without being genotoxic. A wide spectrum of new experimental models should be developed—in vivo models that use animal organs and tissues, cell-culture systems that use human cells and tissue explants, and in vitro systems adequate to permit definition of molecular mechanisms, (e.g., possibly the activation of cellular and viral oncogenes).

The Panel encourages the development of effective short-term tests for the bioassay of environmental chemicals for promoting activity, as an adjunct to those tests that can identify genotoxic initiating agents.

Atherosclerosis is now viewed by some as a proliferative process akin in some respects to cancer. Dietary factors seem to be involved in this process. The exact nature of the mechanisms involved remains to be discovered, but may involve mutation, as well as modulation.

The results of research in these areas are likely to improve our understanding of the differences in modes of action between carcinogens and modulators and may thus form a basis for new approaches to risk estimation and control of chemical exposure.

Thus, it is recommended that a research program be initiated to study differences in modes of action of substances involved in the development of cancer.

VI. DEVELOPMENT OF WAYS TO TAKE THE GREATEST POSSIBLE ADVANTAGE OF EXISTING FEDERAL DATA COLLECTION FOR HUMAN HEALTH EFFECTS ASSESSMENT

NEED: The health impact of environmental chemical exposure of humans is poorly known and largely speculative. Indeed, it may be substantially less than has sometimes been alleged. Reliable estimates require careful epidemiologic studies.

RECOMMENDATION FOR PROMISING RESEARCH: Linkage of existing governmentally collected data sets (such as National Center for Health Statistics, Social Security, and National Institutes of Health data) would provide a useful and efficient basis for epidemiological studies.

ANTICIPATED IMPACT OF RESEARCH: If legitimate concerns over data privacy can be overcome, such linkage would dramatically improve our ability to assess the effects of chemical exposures of humans relatively rapidly and inexpensively.

RATIONALE: There are four events that it is useful to link: exposure, appearance of a preclinical marker, appearance of disease, and death. In the case of many thousands of persons, detailed information has been collected on all four events; but it has been collected at different times, and the sequence has never been linked so as to reveal important patterns. Cost-effective research could be undertaken if the obstacles to such linkage could be removed. For example, within a single federal agency—the National Center for Health Statistics (NCHS)—it would be possible to:

- Identify relationships among occupation, lifestyle, and many environmental characteristics to specific causes of death by linking information obtained on the National Health Interview Survey to deaths identified through the National Death Index (NDI);
- Identify preclinical markers of serious disease by linking information obtained in the Health and Nutrition Examination Survey (HANES) to deaths identified through the NDI or, even better, by re-including some persons in HANES later; and
- Identify preconceptional and pregnancy determinants of disease by linking data from samples of births to deaths of infants identified through the NDI.

The possibilities are further expanded when one considers eventual linkages between data of NCHS and those collected by other federal agencies. Two come quickly to mind: Linkage between NCHS and the Social Security Administration (SSA) to link the SSA 1 percent sample of continuous work histories with deaths ascertained through the NDI, and linkage between NCHS and the National Institutes of Health (NIH) data. NIH, in the course of many very large case-control studies and even larger experimental-intervention studies, has assembled detailed environmental, social, and medical information on many thousands of normal persons (i.e., the controls in the case-control studies, the nonexperimental group in the intervention studies); those data should be followed through the NDI with causes of death linked to the information obtained in the NIH research.

These are only a few of many possibilities. Funding, although needed, is not now the principal constraint. The concern for privacy severely constrains efforts to link data. An approach needs to be developed to determine and protect important aspects of confidentiality—but only those aspects, because the linkages proposed here can provide information of great value to society.

Thus, it is recommended that the policy issues related to data linkage be addressed and that appropriate linkages be made so that existing federal data collection activities can be used to the maximal extent possible to assess the potential human health effects of chemical exposures.

RESEARCH BRIEFING PANEL ON MATERIALS SCIENCE

Theodore H. Geballe ***(Chairman),*** *Stanford University*
Martin Blume, *Brookhaven National Laboratory*
Albert M. Clogston, *Los Alamos Scientific Laboratory*
Billy Lee Crowder, *IBM*
Arthur C. Damask, *Queens College of CUNY, Flushing, New York*
Henry Ehrenreich, *Harvard University*
Fred Gamble, *Exxon Research & Engineering Company*
Stig Hagstrom, *Xerox Corporation*
John K. Hulm, *Westinghouse Research Laboratory*
Herbert Johnson, *Cornell University*
Venkatesh Narayanamurti, *Bell Telephone Laboratories*
William D. Nix, *Stanford University*
Donald R. Uhlmann, *Massachusetts Institute of Technology*
John Wachtman, Jr., *National Bureau of Standards*

Invited Guests

Malcolm R. Beasley, *Stanford University*
Robert L. Byer, *Stanford University*
Robert A. Huggins, *Stanford University*
Lewis H. Nosanow, *National Science Foundation*

Staff

Klaus Zwilsky, ***Executive Director,*** *National Materials Advisory Board*
Donald G. Groves, ***Staff Scientist,*** *National Materials Advisory Board*
Raymond P. Girouard, ***Special Assistant*** *(Stanford University)*
Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON MATERIALS SCIENCE

ABSTRACT

Opportunities exist for extending the frontiers of materials science and for increasing the flow of materials that can lead to new and improved technologies. Materials with the potential for higher performance because of their improved strength, corrosion resistance, and reduced energy consumption are expected to be developed from new types of composites, structural ceramics, and polymers. Materials with the potential for high technology with improved electrical, magnetic, and optical properties are expected from new semiconductor composites, superconducting composites, and optical crystals, glasses, and fibers. Nonequilibrium materials obtained by rapid quenching, surface implantation, and other means are expected to provide a reservoir of technology for reducing problems such as the U.S. dependence on strategic materials. Cooperative phenomena such as solid-liquid transitions, magnetism, superconductivity, metal-insulator transitions, sound waves, charge density waves, and spin waves can be better understood both because of theoretical advances and because of the availability of new materials. The precise description of clean surfaces now available can be extended to technologically important surfaces and provide

direction for producing urgently needed new catalytic materials. The promise of developing the science of fracture now based on macroscopic concepts down into the atomic regime and to be able to make quantitative predictions may be realized. The opportunity to focus on microscale phenomena permeates almost all materials science. The time is ripe for new tools and concepts for exploring phenomena and characterizing materials on a microscale. Instrumentation is an essential ingredient; there is an opportunity to enhance progress in almost all of the areas covered in this report by making state-of-the-art instrumentation and synthesis techniques available to more students and practitioners and by encouraging further development of new methods of synthesis and characterization.

In summary, many areas of materials science offer exciting opportunities for advancing basic science and technology. The Panel is recommending six that seem most promising and timely.

1. Increasing access to the state-of-the-art equipment that is available commercially.
2. Discovering new materials using techniques for synthesis and characterization that depend on physics that is understood but has not been utilized as yet.
3. Making more support and modern equipment available for graduate student training.
4. Investigating phenomena and structures in systems that contain "small" numbers of atoms such as surfaces, interfaces, and grain boundaries.
5. Extending the scientific understanding of the basic science that pervades processing techniques.
6. Developing a science of fracture and other modes of failure based on atomic level models.

INTRODUCTION

Materials science is a broad interdisciplinary research field that brings together the ideas and theoretical understanding and experimental skills of physics, chemistry, metallurgy, and ceramic science. New materials of scientific and technological interest are synthesized in which the mechanical, thermal, chemical, electrical, optical, magnetic, and other physical properties are controlled by understanding the role of defects, impurities, composition, and structure. Among the many past accomplishments of this type of interdisciplinary research is dislocation-free silicon—the building block of silicon-chip technology. An important activity within materials science is the exploration of new materials that may have unusual or unexpected combinations of properties.

Theory and experiment in condensed matter studies are converging as condensed matter theory becomes more proficient at dealing with model systems that approximate real materials in essential ways and as methods of preparing and characterizing materials become more precise. Materials science is well on the way to establishing a fundamental basis for the technology of new engineering materials. The need to optimize more than one property frequently calls for conflicting methods of processing. Establishing acceptable trade-offs is a function that benefits from an understanding of the basic phenomena. The development of its constituent fields to their present level of sophistication has been brought about by both government and industrial support but will increasingly be a major responsibility of government because of the long lead times involved in bringing the basic technology of the complex new materials to commercial viability.

Strategic materials, such as chromium, cobalt, and platinum are largely obtained from abroad, and it is urgent that substitutes for them be found. The most promising and perhaps only nonpolitical route by which alternatives to such strategic materials can be found is through materials science. In general, it can be said that materials are strategic because of the unique properties they provide, such as corrosion resistance, hardness, strength, and oxidation resistance at high temperature and catalytic activity. Higher-priced substitutes exist for some applications, but in others any substitution of present materials results in a loss of performance. However, trends in the development of new high-performance materials promise both cost-effective and/or performance-effective substitution. For example, polymer-matrix materials, glass, and glass-ceramic coatings are promising for corrosion-resistant chemical processing. Surface-alloyed materials offer improved hardness and corrosion resistance. Structural ceramics offer wear resistance and high-temperature service capabilities. Rapidly solidified crystalline metals offer new families of alloys with extended solid solutions, some of which have high strength, hardness, and high-temperature capability. Unpredicted discoveries of new materials can reasonably be expected to occur, and some entirely new solutions to strategic materials problems may appear as a result.

It should be noted from the above considerations that recommendations for research support cannot be simply product oriented. The fundamental approaches to materials science provide the basis for a wide range of applications. For example, particle-solid interaction studies have led to new ion implantation instruments by which accurate diffusion profiles can be measured and semiconductor systems can be made. Without the basic studies, such tools and instruments would not exist—the technology would not exist. New and developing technologies contribute most to the growth of our nation's industrial strength as well as to the generation of new types of employment.

The panel has addressed the opportunities for high-leverage scientific investment in materials science using the following procedure. In this report, classes of materials are identified as high technology for utilization in electronic, optical, magnetic, and acoustical devices and as high performance for utilization in mechanical, chemical, or thermal devices. The fundamental properties of these materials are described, and the basic science required for increasing the understanding, discussed. Underlying the research is the need for the accurate characterization of materials. Opportunities for new instrument development are illustrated.

Materials science, with its multiple roots and many-branched interfaces, is less monolithic than, say, astronomy or particle physics; consequently, obtaining a consensus of research priorities is less straightforward. Much research involves relatively small numbers of investigators. As a result, there can be rapid changes in response to new stimuli so that worked-out fields are rapidly abandoned.

Choices must be made between emphasis on the large national facilities versus small institutional-based research. Both are essential, but the balance must be continuously monitored. The ability to develop and bring large-scale facilities to fruition must be maintained, even at the expense of phasing out or reducing the scope of still-valuable facilities that are no longer competitive in producing exciting new science.

Four national synchrotron radiation-source facilities are now in operation and serve many different constituencies in the scientific, medical, and industrial sectors. The new science and advances in instrumentation, characterization, and technology that have been accomplished, and those that should be forthcoming, justify, at present, financial support at something close to the projected levels.

Neutron spectroscopic and radiation facilities exist at national laboratories (Brookhaven, Oak Ridge, Los Alamos, Argonne) and other government laboratories. The problem of how to set priorities and take advantage of new advances has already been addressed by the Review Panel on Neutron Scattering (Brinkman) committee. The cost-effectiveness of national facilities such as the Magnet Laboratory at MIT, the submicron facility at Cornell, the Solar Energy Research Institute, and other special-purpose laboratories should be evaluated with respect to the totality of research and technology being done by smaller institution-based groups. In general it is believed that the most effective use of resources is to build on the expertise at existing laboratories rather than to create new ones. The training of workers with state-of-the-art equipment of new research must be an important consideration in order to overcome the present-day shortage of good materials scientists in industry.

HIGH-PERFORMANCE STRUCTURAL AND SPECIAL-PURPOSE MATERIALS

These are materials whose intrinsic characteristics, which are determined by their components and the methods by which they are synthesized, enable them to give improved performance.

COMPOSITES AND STRUCTURAL CERAMICS

Use of improved composites and structural ceramics can increase energy efficiency; reduce dependence on strategic materials; and help keep U.S. products such as aircraft, automobiles, and power turbines competitive with foreign ones.

Promising composites include high-modulus fibers (such as graphite or silicon carbide) in several types of matrices, including polymers, metals, and ceramics. Higher strength-to-weight ratios, high toughness, improved fatigue resistance, and high corrosion resistance can be achieved. *Realization depends on improved interfacial characterization and control and on processing science to achieve optimum distribution of the components.* Advances in spectroscopy applied to interface characterization and in the rheology of fiber/liquid suspensions give promise of providing a science base to support processing developments.

Promising structural ceramics include silicon carbide, silicon nitride, and related materials. Sufficiently high levels of strength and toughness already achieved permit short-term (100-hour) application up to 1100°C. Applications requiring a longer service life and higher temperatures appear possible if stronger grain boundaries can be achieved. Use of these improved materials can increase energy efficiency in diesel and turbine engines and extend service life in many types of machinery that are subject to high wear rates. Grain boundaries in the present material have relatively poor long-term strength and oxidation resistance because of additives needed for sintering. Improved powder production and control of mass transport in the boundaries appears possible. Modern surface characterization methods and analytical electron microscopy offer the opportunity to place the design of grain boundaries (including impurity concentration effects) on a scientific basis to guide optimization of sintering with reduced levels of long-term creep and oxidation.

Another important class of new structural ceramics is the family of transformation-toughened ceramics typified by alumina strengthened by zirconia inclusions. Significant increases in strength and toughness have been achieved. The technique is promising for strengthening other ceramics, but fundamental scientific questions concerning the mechanism of strengthening need to be answered. Several theoretical models exist; detailed experiments involving structural characterization on a microscale are required and appear possible.

POLYMERS

The use of polymers has increased by 10 to 15 percent each year in the last 50 years—the fastest growth rate of any class of materials. Polymers are now used, or are considered for use, as critical components in materials systems ranging from structural composites and ophthalmic lenses to semiconductor packaging and chip fabrication.

Among the areas where future developments are limited by a lack of scientific understanding, and where improved understanding could have great impact on technology, four are particularly noteworthy: (1) liquid-crystal polymers, (2) high-temperature polymers, (3) lithographic polymers, and (4) polymeric adhesives.

Liquid-crystal (rigid-chain) polymers offer exceptional, and previously unobtainable, combinations of properties (a tensile strength approaching that of steel yet with a density less than 1.5 g/cm^{-3} and relatively high thermal conductivity and low thermal expansion) whose applications range from reinforcing fibers to bulk moldable materials, from aircraft with larger payloads to automobiles with greater fuel efficiency. The effective development of this relatively unexplored class of polymers depends critically, however, on acquiring *improved understanding of the statistical physics of rigid-chain molecules and of the* *structure, rheology, crystallization, and relaxation behavior of the materials.*

High-temperature polymers (polymers with use temperatures exceeding 300°C) offer outstanding opportunity in areas ranging from novel protective coatings to semiconductor packaging. The achievement of these opportunities depends on an imaginative effort in synthesis, processing, and characterization. Promising chemical approaches such as combining organometallic catalysts with polymer substrates are largely unexplored, as are the chemical and structural changes that take place on curing and the effects of these changes on mechanical, thermal, and chemical properties.

Lithographic polymers are essential elements in the fabrication of integrated circuit devices. The demands of large-scale integration and production efficiency require new polymers with increased sensitivity, resolution, and difference in etch rates between exposed and unexposed regions. These demands require a better understanding of ultraviolet, x-ray, and electron-beam exposure and, in turn, of polymers with the requisite sensitivity to these radiations. Improved polymers in this area would provide great leverage for integrated circuit manufacturers; yet only small volumes of materials are required, and this provides reduced incentive for industrial groups with traditional polymer expertise. *Future developments in lithographic polymers will depend on fundamental investigations of the interaction of various radiations with polymers, synthesis of novel polymers with tailored combinations of properties, novel approaches to the resist function (as polymers that fall apart and vaporize on exposure to appropriate radiation), and understanding of relevant processes such as reactive ion etching.*

Novel polymeric adhesives offer the promise of effecting radical changes in manufacturing technology and product design. At the same time, *developments of the past decade in surface spectroscopic techniques and fracture mechanics offer the possibility of obtaining critical new insight into the nature of adhesion,* its sensitivity to environmental factors, the fracture mechanics of adhesive joints, and the role of coupling agents and interphase material. Work in these areas, dealing with polymers whose chemical and physical structure is tailor-made, should dramatically change current technology. The need for understanding is critical; the tools for providing that understanding already exist.

High-performance materials are frequently subject to high material stresses in service. The failure of such materials can be quite costly or even catastrophic. Much effort and cost go into ensuring safe and reliable service. Opportunities exist to increase reliability and reduce costs by coupling improved understanding of failure processes with improved diagnostic techniques.

HIGH-TECHNOLOGY ELECTRICAL, MAGNETIC, ELECTRONIC, AND OPTICAL MATERIALS

These materials are the basic building blocks of the electric power, control, computer, and communications industries. Advances in electronics, in particular microcircuits, have been so rapid in the past two decades that they are producing a major revolution in information technology. These advances are based on materials-processing technology at a sophisticated level where complex microcircuit structures composed of many layers of different thin films can be fabricated with precise control on a scale of 1 micron or less. The materials technologies discussed below are of growing importance.

ELECTRICAL AND MAGNETIC MATERIALS

The power generation, distribution, and control industries already rely heavily on sophisticated magnetic alloys and advanced composite polymer insulators. New materials that are of growing importance in electrical applications include large area silicon crystals for power electrical devices, which play a rapidly growing role in AC to DC power conversion and variable frequency devices. Advanced materials under development include high-field high-current superconducting composites for making electrical generators more efficient. Improved multifilamentary superconducting composites are essential for providing magnetic confinement in nuclear fusion plants. High saturation, low-loss amorphous magnetic materials have technological promise for transformers and machines.

ELECTRONIC MATERIALS

Beginning with basic work at Bell Telephone Laboratories on the junction transistor in 1948, a new materials-based solid-state electronic technology has developed in the past 30 years that has produced staggering advances in the arts of communication, computers, command, and control. The first wave of this new technology was based on the ability to grow defect-free, high-purity single crystals of silicon and the ability to fabricate micron-scale devices with metal-oxide-semiconductor layers arranged in microscale transistor circuits. The fabrication of microcircuits represents probably the most advanced materials processing developed to date. The evidence suggests that we have only scratched the surface of the opportunities in this area and that enormous growth is still to come. The new electronic technologies now emerging will have major influence on the entire spectrum of U.S. industry—in communications, automation, command, and control.

These new technologies will utilize not only silicon but also semiconductor crystals such as gallium arsenide, indium phosphide, mercury cadmium telluride, and silicon carbide. Integrated circuits will be coupled with advanced sensors to detect signals over the entire electromagnetic spectrum. Signals will be processed at higher and higher speeds and will involve cryogenic semiconductor and superconducting Josephson-junction thin-layer devices. They will use other, nonelectromagnetic methods such as surface acoustic wave techniques. Each of these requires a multitude of new electronic crystals, new thin-film and multiple-layer materials and processing methods using photons, coherent photons, electron beams, x-ray, and other advanced processing techniques. Magnetic films with vertical anisotropy are generating excitement because of their greatly increased digital storage capacity. Amorphous rare earth-based materials may be the long-sought medium for magnetooptical recording.

OPTICAL MATERIALS

Running close to the electronic revolution is a new era of optical technology that ultimately will be the main channel for most information transmission in the world. This will involve national broadcast, cable and telephone signals, computer networks, and control circuits in industrial plants. This new communication medium is being made possible by three independent but converging materials developments:

1. Low-loss optical fibers.
2. Lasers, in particular micro-sized solid-state lasers.
3. Solid-state photon detectors and electrooptical devices.

Rapid advances in optical materials and electrooptical devices are expected in the next 20 years, which may lead to other revolutionary changes in process control. A plethora of new electrooptical, acoustooptical, and other cross-coupled materials are emerging, such as (optical) damage-resistant lithium niobate and thallium arsenic selenide. Single crystals will be required in most cases and will require new advanced methods of growing high-purity, defect-free single crystals, such as single-crystal fibers, which can be used for remote sensing. This general field will greatly influence avionics and defense technology and represents one of the greatest U.S. technological opportunities on the horizon.

Advances in glass (Nd-doped) and crystal (garnet) laser materials can be expected to double and even triple the efficiency of the lasers in the next 10 years. Even more important than increased efficiency might be the reduced size and weight that will permit the mounting of multikilowatt lasers on robot arms for such applications as cutting, drilling, and welding.

NONEQUILIBRIUM MATERIALS

Materials far from equilibrium are typically produced by processes that prevent lower free-energy or ground-state structures and compositions from forming. A particularly valuable method for controlling the kinetics is by quenching from the melt and the vapor. A remarkable variety of nonequilibrium structures can be produced, with resulting properties that are of great scientific and technological interest, by varying the processing conditions. *Opportunities exist in both the understanding and the control of the material transformations that occur in these processes.* Spectroscopic techniques for very rapid measurement of local temperatures, reaction rates, and mass transport of chemical species, including short-lived intermediate species, can assist in building a science base for processing control and for suggesting process modification. Opportunities exist for improving the characterization of the novel structures produced. These include new glasses, extended ranges of solid solutions, gradient structures, layer structures, and modified surfaces. An emerging field of materials science concerns the relation of these novel structures to unique and desirable properties on the one hand and to the process details required for producing these structures on the other.

The existence of new kinds of materials and processes may provide a way to reduce our nation's dependence on certain strategic elements. For example, *rapid solidification methods provide ways of creating alloys with entirely new chemical compositions,* compositions that could not be explored previously because of the limitations of conventional processing. The addition of transition elements such as

Fe and Mn to aluminum alloys for increasing elevated temperature strength is now possible because of this development. Thus, a *much broader range of opportunities now exists for selecting more abundant elements for the synthesis of high-performance materials.*

Processing of surfaces by laser heating can be used to improve the wear and corrosion resistance of metals significantly. The strategic material chromium, which is required for corrosion resistance in stainless steels, is needed primarily at the surface where the corrosion takes place. Through the use of new surface processing methods it might be possible to create stainless steels using but a small fraction of the chromium now required for this application. Ion implantation opens the way to forming new surface alloys. These can be crack resistant, corrosion resistant, and wear resistant. Other alloys might have catalytic properties.

Substitutional nitrogen, an important impurity for the basic understanding of deep-level defects in semiconductors, was first produced in silicon by ion implantation and laser annealing after other techniques failed. This technique has recently been used to enhance by a factor of 10^5 the photosensitivity of transparent ferroelectric ceramics and to cause significant changes in the spectral sensitivity. These new materials find use as nonvolatile and electrically erasable image-storage devices.

OPPORTUNITIES IN CONDENSED MATTER SCIENCE

Nonequilibrium structures provide a powerful approach to the testing of our basic understanding of condensed matter. The extended compositions available allow theories of physical, chemical, and in some instances mechanical properties to be evaluated over much wider ranges of the relevant variables. For example, important yet subtle concepts of the role of structural (atomic) order/disorder may be tested through studies of amorphous and crystalline materials of the same chemical composition.

SOLID-STATE COOPERATIVE PHENOMENA

In a fundamental sense, all solids and liquids exhibit properties based on the interactions of very large numbers of electrons and atomic nuclei whether these are best represented as neutral atoms, charged ions, or mobile electrons moving within a lattice of fixed positive ions. The range of phenomena encountered in such complex, many-body systems is great. In particular, such systems exhibit emergent, cooperative phenomena that are not present in disassembled arrays of separate atoms. The simplest such phenomenon is the crystallization of solids into regular arrays of atoms or ions. Others are the thermodynamic transitions between different solid and liquid phases, magnetism, superconductivity, metal-insulator transitions, sound waves, charge waves, and spin waves. As mentioned below, some particularly important cooperative phenomena occur in materials systems that are effectively one- and two-dimensional. *Since the site of action of many chemical and electronic phenomena, such as catalysis and transistor action, is on surfaces, the discovery of cooperative phenomena in two-dimensional systems opens exciting scientific and technological opportunities.*

The difficulties of dealing theoretically and experimentally with cooperative phenomena in condensed materials led physicists at first to seek out the simplest systems that might be understood on an atomic basis, such as the pure monovalent metals and pure single crystal semiconductors. In these cases useful approximations could be made based on the motion of single electrons, and cooperative phenomena did not play a prominent role. At this earlier time, a fundamental atomic understanding of the stabilization of crystal phase, magnetism, superconductivity, and other cooperative phenomena was beyond the explanatory power of condensed matter physics.

Today, because of the development of more powerful theoretical and experimental methods, the situation is quite different. The ground state energy of ordered crystal phases can be calculated in many cases from first principles to high numerical accuracy using methods that take account of the many-body interactions. Matter is studied under extreme conditions of high pressures and low temperatures where new thermodynamic states and transitions can be found to test new ideas. Second-order phase changes are well understood because of the introduction of renormalization group methods. Superconductivity is better understood on an atomic basis as a result of a series of theoretical developments starting with the Bardeen-Cooper-Schrieffer (BCS) theory but is not yet capable of predicting transition temperatures or explaining why critical temperatures above 23 K have not been found experimentally. Magnetism is now much better understood both in terms of the atomic interactions giving rise to spin alignment and the phase transitions between magnetic and nonmagnetic states. Two other cooperative phenomena closely related to magnetism are now beginning to be well understood. The first is valence fluctuation whereby rare earth atoms in certain compounds do not possess a single fixed valence but fluctuate dynamically between two charge states. The second is the metal-insulator transitions in which transition metal oxide and some other materials exhibit a transition temperature at which the compound changes its conductivity by many orders of magnitude from a conducting to an insulating state. *Both of these phenomena share with magnetism the fact that they are bulk properties of matter and therefore can have large macroscopic effects. They have not yet found a firm place in materials technology but deserve further attention because of their potential for useful applications.*

It is a fact of quantum physics that systems exist in nature that effectively exhibit one- and two-dimensional behavior and have many practical effects. The basic reason is that a limitation on the transverse motion of a quantum mechanical system, because of zero point effects, makes the system effectively one- or two-dimensional for small enough excitations. The most exciting recent developments in cooperative phenomena have been in reduced dimensionality systems. In two dimensions, for example, strong cooperative effects have been found in semiconductor inversion layers. In the presence of a magnetic field, two-dimensional inversion layers on Si and GaAs show a totally unexpected, accurately quantized Hall effect that is leading to an improved determination of the ratio of the fundamental constants e/h. In a two-dimensional system comprised of electrons trapped on the surface of liquid helium, Wigner crystallization has been observed for the first time, whereby a charged plasma of electrons condenses into a two-dimensional electron crystal. In other two-dimensional systems, such as monolayer liquid crystal films and layers of atoms intercalated into layered transition metal disulphides, the melting of two-dimensional crystals has been observed.

Another partially understood cooperative phenomenon is that of *localization* of electron motion in disordered solids. This is an effect whereby an electron of sufficiently low kinetic energy may actually become trapped in a localized region of a disordered conductor. The effect, particularly marked in one- and two-dimensional systems, leads to strong departures from ohmic conduction in partial agreement with theory. This effect could become one of practical significance in large-scale silicon integrated circuits as the characteristic dimensions of conduction paths shrink below 0.1 micron, particularly if integrated circuits of the future are operated at cryogenic temperatures to increase electron mobility and improve heat transfer.

Another class of one-dimensional cooperative phenomena arises in the conducting polymers such as polyacetylene where conduction takes place by electron transfer along linear chains of molecules or through linear stacks of molecules such as in the organic conductor tetrathiafulvalene-tetracyanoquinodimethane (TTF TCNQ). These materials show many cooperative phenomena such as spin waves, charge waves, and the so-called Peierls instability. In some cases they show soliton formation whereby a coupled electron-nuclear system develops a high amplitude motion that propagates as a charged entity with many of the properties of a free noninteracting particle.

On the very forefront of the physics of cooperative phenomena is a poor but growing understanding of the way in which nonlinear effects in systems far from equilibrium showing charge or mass transport can lead to bifurcation of their periodic response followed by the onset of chaos and turbulence as the driving field is increased. Examples of such instabilities occur in hydrodynamic flow, chemical reactions, phase transitions, and nonlinear electronic circuits.

Cooperative phenomena, as the above examples show, generally tend to arise, or find their strongest expression, in the more complex material systems such as high-temperature superconductors (the A15 structures), bubble memory ferromagnets (the yttrium iron garnets and analogs), the quantized Hall effect (artifically layered GaAs), the metal-insulator transition (vanadium oxides), two-dimensional melting (liquid crystals, monolayers of rare gases), and solition formation (polyacetylene). Research efforts form strong links between condensed matter physics, metallurgy, and solid-state chemistry and lead to the convergence of these fields into a unified materials science.

SURFACE SCIENCE

The study of clean surfaces is an active and rapidly progressing field that moves hand in hand with ultra-high vacuum technology. The structure of the different faces of metals and semiconductor crystals can be determined in great detail both theoretically and experimentally. The calculated displacements of the outer atoms from their unrelaxed position (i.e., a position determined by an integral number of lattice translations from the interior) are often in good agreement with the reconstructed surface symmetry observed by low-energy electron diffraction and with the atomic positions determined by angular-resolved photoelectron spectroscopy and other surface-sensitive techniques. The positions and bonding of ad-atoms can also be calculated and observed. Field ion microscopy can be used to follow the motion of the ad-atoms. There is an opportunity for extending the studies to many more surfaces, including technologically important ones, to learn more about the behavior of ad-atoms and to move in the direction of being able to understand the atomic behavior of surfaces involved in chemical reactions.

Surface chemistry includes such areas as catalysis, in which chemical reactions take place on surfaces having highly controlled chemical and structural properties; corrosion, in which undesirable modifications of materials properties are generated by surface chemical reactions; and photochemical surface reactions, which offer promise of new ways to synthesize chemicals.

The phenomena involved in corrosion are of technological importance in areas ranging from buildings and boats to nuclear waste disposal and high-performance batteries. Recent advances in surface spectroscopic techniques have opened the possibility of obtaining new insight into the underlying chemistry and physics and the opportunity of developing more detailed models to describe corrosion phenomena. The role of corrosion inhibitors can also be explored in depth, and this in turn should lead to descriptions of the relevant thermodynamics and kinetics.

The area of adhesion provides another example of surface-dominated phenomena. The ability to form adhesive bonds that are stable in hostile environments would be greatly facilitated by improved understanding of the chemistry, structure, and mechanics of interfaces. Advances in surface spectroscopic techniques and in techniques for chemical modification of surfaces offer outstanding opportunities for developing new insight.

Heterogeneous catalysis is a field based on solid materials whose surface chemistry enables them to facilitate the transformation of less valuable molecules into more valuable ones. Heterogeneous catalysts are at the heart of the petroleum, petrochemical, and chemical industries. As such, they are usually thought to lie in the domain of chemical engineering and are not considered a part of materials science. This state of affairs, while understandable, is not a reflection of the nature of the situation. *There is an opportunity for extending surface characterization techniques to the study of the kinetics of surface-catalyzed reactions.*

In the petroleum and petrochemical area, the shift is to feedstocks that are heavier (larger molecules) and dirtier (lots of S, N, etc.). Crude-oil quality is decreasing. Improved catalysts are needed to reduce the size of molecules as well as their S and N contents. In addition, while the synthetic fuels era is further off than was thought a few years ago, it is only delayed 10 years or so from earlier scenarios. This extra breathing time can be used to improve our technology. For example, shale oil contains a higher concentration of nitrogen than does ordinary petroleum. This must be removed. Catalysts are sought that will permit the removal of nitrogen selectively, i.e., without the consumption of excessive hydrogen.

One of the largest technologically important surface problems is associated with the internal surfaces of the porous rocks from which we extract oil. Oil reservoirs are formations of porous rock of enormous surface area filled with oil. When oil is produced, only a fraction of it, less than half, comes from the rock. The balance is held in place, absorbed on the rock's internal surfaces. Thus, we have in place in the ground in abandoned fields oil "reserves" that exceed all that have been consumed to date.

An underutilized energy resource is natural gas in remote locations. Economic processes need to be developed that can convert methane into liquids. A catalyst that could convert methane directly to such products does not exist. If it did, the economics for such a process would improve dramatically.

Syn-gas or CO/H_2 is going to be a major feedstock of the future. It will be the product of coal gasification and from it much of our fuels and petrochemicals may be derived. Today catalysts do not exist that can take CO/H_2 selectively to desired products. Instead, a spectrum that must be separated and processed further is obtained.

STRENGTH AND FRACTURE PHENOMENA

Ultimate strengths of materials are known from theory to be very high and have actually been observed to be so in a few defect-free, structurally perfect materials. *There is a clear need for further study of the fracture physics of the newer high-strength materials,* especially fibers such as graphite and high modulus polymers such as Kevlar. The fundamental factors controlling fracture of these materials, and their relation to process parameters, environmental conditions, etc., are not well understood.

Failure of most practical materials occurs at much lower stresses, because of a multiplicity of factors, including stress concentrations, the motion and interaction of defects, and environmental factors. The role of defects has been conceptually clear for many years, and some progress has been made in treating the behavior of individual dislocations, cracks, and vacancies. The latter are of special importance in high-temperature fracture.

The extreme mathematical complexity of dealing with the interacting systems of defects involved in deformation and fracture has largely prevented the application of fundamental theory to practical conditions. The situation is now changing as a result of recent studies on both the atomic and continuum levels. On the atomistic level, modern computational methods have made it possible to treat defect interactions and to extend such models to large aggregates of defects.

Several recent developments offer the promise of leading to a true science of fracture based on atomic-level models. A realistic two-dimensional model of the atomic structure of a crack has finally been solved in closed form and used to describe a form of slow fracture associated with "lattice trapping." At the same time new developments in continuum descriptions of the near-crack region are being used in experimental studies of crazing in polymers. On yet another front, a detailed chemical model has been used to predict which polar molecules are potent stress-corrosion agents for silicates.

The way now seems open to combine discrete lattice effects, interacting defects, molecular models of chemical reactivity, and continuous mechanism treatments in a stress field to produce a fundamental theory of fracture that could be quantitatively tested against experiment. A relatively modest increase in support might well have high leverage.

SCIENTIFIC BASIS OF SYNTHESIS AND PROCESSING

During the last 20 years there has been a rapid growth of synthesis methods performed in a vacuum. Examples of such techniques are evaporation, sputtering, chemical vapor deposition (CVD), and plasma deposition. For many of these techniques the basic physical mechanisms behind the deposition are still not well understood—for example, the sputtering process or the plasma deposition. Two techniques look particularly promising for the epitaxial growth of semiconductor materials with a much larger field of applications. These are molecular beam epitaxy (MBE) and metallo-organic chemical vapor deposition (MO-CVD).

The development of ultra-high vacuum (UHV) technology ($< 10^{-10}$ torr) has allowed the development of MBE as a method for the synthesis of structures by controlling the deposition of each layer as it is being deposited. The broader application of UHV technology to materials synthesis will allow fabrication of high-purity, better-controlled materials. This approach is thought to be far more interesting and promising than materials synthesis under reduced gravitational gradients (space synthesis).

The other epitaxy growth technique that seems particularly interesting is MO-CVD. Here the growth occurs by the decomposition of metallo-organic molecules in a thermal reactor. The technique allows a close control of the growth by monitoring the gas flows into the reaction chamber. It provides exceptionally high quality semiconductor materials, especially in the III-IV compounds family.

New process development is often a quasi-empirical activity. Processes are inherently complex; the important parameters are usually numerous and sometimes not well defined. The scientific base for the various stages of a "typical process" is often inadequate. Consequently, process development may be a frustrating and time-consuming activity.

The science of processing would be greatly advanced by increased understanding of some basic phenomena that pervade many processing techniques. Nucleation is an important example. Nucleation of a new phase is an important aspect of virtually all processing techniques, microscopic and macroscopic. Processes that would benefit from a better understanding of nucleation include rapid solidification, vapor deposition, production of ultra-fine and monodisperse powders, and sol/gel methods. Lack of understanding of transport across and along interfaces is another area that limits process development. New studies of mass and heat transport across interfaces, often with unconventional geometries and boundary conditions, would facilitate process development. In particular we note the opportunities in ion beam technologies with the use of ion accelerators for both materials characterization and materials synthesis/modification (ion implantation) and in inelastic scattering of neutral atoms.

Processes based on sol/gel technology offer promise for the production of novel glasses and ceramics—materials with novel microstructures and properties. Developments in this area depend on improved understanding of the transitions: sol-gel-dried, gel-glass-crystalline ceramic, of the conditions necessary for the formation of gels (including the effects of pH and anions), and of the structure and properties of gel-derived materials. Also of promise are mixed organic-inorganic materials derived from gels, although not enough is known about their structure and properties.

Techniques for producing monodisperse fine powders depend on improved understanding of nucleation kinetics (in the gas phase as well as in condensed phases), of the growth and coalescence kinetics of small (500–5000 Å) particles, of the interactions between such particles, and of the phenomena involved in comminution.

Conversion of particles or dense bodies depends on improved understanding of particle-binder interactions, of drying phenomena, of sintering kinetics (particularly the sintering kinetics of covalently bonded materials), and of the role of dopants as sintering aids or grain growth inhibitors. Exceptional processes for enhanced dimensional control include injection molding of ceramics and casting of polymer-matrix composites that display no volume change on cure. Developments in this area depend on improved understanding of the rheology of highly loaded suspensions, of the dynamics of injection molding for such suspensions, and of the dependence of volume change during polymerization on chemical structure.

There is a great need for fundamental studies to elucidate the chemical and physical structures of polymeric materials. Controversy exists concerning the basic microstructural units of semicrystalline polymers (adjacent re-entry folding versus switchboard models for the lamellar plates); reptation and (snake wiggling) center-of-mass diffusion are both strongly advocated as the dominant processes for diffusive transport in the liquid state; important structural features of liquid crystal polymers remain largely uncharacterized; the chemical and structural characteristics of thermosetting polymers, and their dependence on process history, are substantially unexplored; and the detailed roles of such modifiers as rubbery second-phase material in semi-crystalline polymers have not been clarified. Work in areas such as these should yield improvements in our understanding of and our ability to modify the properties of the materials.

INSTRUMENTATION

Access to modern instrumentation is a prerequisite for progress in nearly all areas of materials. In the cycle, new materials synthesis/processing/characterization/application, the development and use of new instruments has allowed a shortening of the cycle time and the development of completely new classes of materials. The needs of the materials scientist in this cycle have formed an important impetus for the development of new and sophisticated instruments. Discussed below are some aspects that have had a profound influence on the development of instrumentation for materials science and where it is expected that further investments will continue to rapidly advance the state of the art.

Experience with state-of-the-art equipment is essential for the training of materials scientists. Major investments in instrumentation were perceived to be necessary by the five industrial members of the Panel to meet the need for well-qualified Ph.D. graduates as well as for new scientific discoveries.

CHARACTERIZATION TECHNIQUES

The development of ultra high vacuum technology has also promoted progress in surface science and surface analysis. By being able to keep a surface clean for several hours it is possible to make a thorough characterization of that surface and to do experiments on it. The extension of UHV technology to other characterization techniques, e.g., electron microscopy, is just emerging. The possibility of combining surface analytical tools such as Auger and photoelectron spectroscopy with electron microscopy and energy loss spectroscopy in an UHV environment is an instrumentation opportunity with high leverage now being realized at a few institutions.

Many materials characterization techniques are based on beam spectroscopies. A beam of photons or particles impinges on the sample, and outgoing photons or particles are analyzed with respect to energy, intensity, polarization, and angular distribution. Progress in condensed matter physics in the understanding of beam interactions with matter has promoted the use of these techniques as analytical tools. There are many opportunities for further progress in this field with a strong coupling between the physics and the materials science communities.

The further development of new sources of electromagnetic radiation such as synchrotron radiation sources, high-power lasers, and free-electron lasers provide major opportunities for applications in materials science. The rapid development of surface-sensitive analytical techniques will be further enhanced by access to these sources.

The sensitivity of some of these techniques has been pushed to a spectacular degree. Trace impurities in microsized regions at the level of 1 part in 10^8 can be detected in favorable cases. In sharp contrast to these advances stands the fact that we are still looking for good analytical tools for the analysis of light elements. Surface analysis has advanced to where fractions of a monolayer of elements heavier than helium can be detected. However, the detection of light elements even in the percentage range becomes difficult when they are embedded in the bulk. The situation appears to be particularly cumbersome for insulators and organic-based materials systems.

NATIONAL FACILITIES

The generation of some of the beams mentioned previously requires large facilities, such as nuclear reactors or accelerators for neutron beams, electron storage rings for synchrotron radiation, and high-voltage facilities for electron microscopy. Continued optimal use of these facilities for materials science depends on continued instrumentation development.

Synchrotron radiation sources provide such capabilities as small angle scattering, crystallography, topography, x-ray microscopes, microprobes, EXAFS, and photoemission spectroscopy. These can be extended with the development of devices such as wigglers and undulators that make it possible to study time-dependent phenomena. Some of the necessary instrumentation developments are made possible by progress in materials fabrication such as artificially structured x-ray monochromators and x-ray transmission optics.

Neutron sources provide unique capabilities for the study of materials. Magnetic scattering, deuterium substitution, inelastic scattering, and small-angle scattering are complemented by similar x-ray methods. Instrumentation advances will provide new methods for high-energy resolution of particular importance in the study of polymers, for example.

It is obvious that major national facilities must serve wide geographical areas. However, even more modest instrumentation facilities like electron microscopy, NMR, surface analysis, and ion beam accelerators have reached such levels in cost and operation sophistication that a certain centralization is necessary. The National Science Foundation's regional instrumentation centers are experiments to try to solve some of these problems, and they should be watched closely to allow an evaluation of the success of this approach. Similarly, the materials research laboratories serve an important function in this respect on a campus basis.

INTEGRATED MATERIALS PREPARATION AND CHARACTERIZATION SYSTEMS

The ultimate coupling between materials synthesis and materials characterization steps is achieved by combining the two in the same system, allowing materials characterization to be done simultaneously with the materials growth as is done, for example, in MBE growth. This trend will be further enhanced with the widespread use of microcomputer-controlled equipment and characterization tools. An opportunity exists in the development of appropriate sensors to monitor the process parameters.

COMPUTERS

Access to powerful computers will continue to be a prerequisite for further progress in the theoretical understanding of materials. It is interesting to note that today it is possible to compute the ground-state energy of a given structure and to predict the phase diagram of a certain multicomponent systems. Access to large computers is also a necessity for the successful modeling of materials systems and processes. The ultimate goal is to obtain such a good model of a material system that an accurate prediction of its properties can be made before it is produced.

As a result of the electronic revolution (partly enabled by the spectacular advances in materials science), we have witnessed a dramatic decrease in computational cost per operation and the development of powerful computers on a chip basis. With access to modern custom design and fabrication tools it is even possible to make "custom" computers in the form of special chips designed for special types of computation, all within reasonable cost limits. A rapid assimilation of this technology applied to materials science represents a unique opportunity.

Computers in the form of microprocessors will play an increasingly important role in the data acquisition and process control of materials fabrication. The rapid growth of computer networks will allow remotely controlled experiments at the large facilities. These networks will play a major role in the dissemination of information as well as in allowing access to powerful information database systems. A challenging task is the training of materials scientists in the use of modern electronics and computers.

ANALYSIS OF HAZARDOUS MATERIALS

Although the materials science issues of hazardous materials are not examined in this report, a few comments on their analysis are in order. Safety levels have moved downward as more sensitive analytical tools have become available.

With the advances in sensor technology coupled with microprocessors and data links, it is now possible to monitor very closely potentially hazardous environments and to be able to take appropriate measures to avoid accidents and catastrophes. There remains a strong need for rapid materials analysis and evaluation in awkward field situations, such as a chemical spill on a freeway.

STATE-OF-THE-ART INSTRUMENTATION

The successful development of instrumentation requires a strong coupling of the application of physical phenomena; the use of modern measurement technology; and the needs of, for example, the materials scientist. An opportunity exists in creating an environment that increases this coupling. To wait until commercial instrumentation becomes available will often lead to a missed opportunity. This coupling usually occurs naturally in an industrial research laboratory environment but to a considerably lesser degree at American universities.

COST OF INSTRUMENTATION

With the development of more sophisticated instruments and because of the higher rate of inflation for instruments, the cost of instrumentation has increased sharply. This has posed a serious problem, especially in universities, in keeping up with instrumentation development. A partial solution to this problem is the use of centralized facilities on a regional or campus basis. However, progress in materials science today requires a substantial instrumentation budget for each individual scientist.

CONCLUSION

Opportunities in materials science have been discussed from the viewpoints of the technologies involved, the new materials and processes that can be foreseen, and the challenges to the basic sciences and engineering disciplines that constitute materials science.

Certain themes are pervasive. A better understanding of surfaces, interfaces, and grain boundaries can be achieved, which will lead to more control over synthesis and to new materials with improved properties. This in turn will affect all processes dependent on strategic materials and on the structural, power, and transportation industries. It will lead to new, faster, and cheaper microdevices for the data processing and communications industries. Much of the progress foreseen depends on realizing opportunities for improving instrumentation and its attendant technology.

Concepts and models being pursued in the basic science underpinning of materials science give intellectual stimulation, provide excellent university-based research projects for the training of the graduate students needed by industry, and occasionally provide entirely unforeseen innovations.

THE REPORTS OF THE RESEARCH BRIEFING PANELS

1983

RESEARCH BRIEFING PANEL ON THE SOLID EARTH SCIENCES

Charles L. Drake ***(Co-chairman),*** *Dartmouth College*

Don L. Anderson ***(Co-chairman),*** *California Institute of Technology*

William R. Dickinson, *University of Arizona*

Carl Kisslinger, *University of Colorado*

John C. Maxwell, *University of Texas at Austin*

V. Rama Murthy, *University of Minnesota*

Jack E. Oliver, *Cornell University*

C. Barry Raleigh, *Lamont-Doherty Geological Observatory*

Frank M. Richter, *University of Chicago*

Eugene M. Shoemaker, *U.S. Geological Survey*

Edward M. Stolper, *California Institute of Technology*

Peter J. Wyllie, *California Institute of Technology*

Staff

Pembroke J. Hart, ***Executive Secretary,*** *Geophysics Research Forum*

Barbara Valentino, ***Staff Associate,*** *Geophysics Research Forum*

Patricia Miron, ***Secretary***

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON THE SOLID EARTH SCIENCES

SUMMARY

The Panel on Solid Earth Sciences has selected five broad targets of opportunity that have the capacity to provide rapid advances in the near future and that will contribute most to our understanding of the Earth's interior and history. The common themes are the structure, composition and evolution of the continental lithosphere, and the dynamics of tectonic processes. This report emphasizes areas that lie on the frontiers of earth sciences—in some cases on the frontiers of chemistry and materials sciences as well—and describes some of the conceptual and technical advances that make it possible to explore more fully the third and fourth dimensions, depth and time. With additional resources, available skills and techniques can be applied immediately to some of these basic problems. Some projects, such as deep continental drilling, determination of the continental geoid, and crustal seismic reflection, can proceed immediately if resources are made available. Others, such as large seismic arrays, expanded isotopic exploration of the crust and mantle, monitoring of crustal motions, and the chemistry and physics of geological materials, require major investments in modern facilities.

With modern facilities and instrumentation the basic research opportunities of these projects will provide excellent training for the next generation of scientific leadership in the earth sciences as well as the advanced skills and trained personnel sought by a wide variety of industries, especially those seeking natural material and energy resources.

INTRODUCTION

Twenty years ago the hypothesis of sea-floor spreading was proposed. It led to general acceptance of the concept of a dynamic Earth and to the plate tectonics model as an explanation for the principal tectonic features. In this model the rigid outer shell of the Earth, the lithosphere, is broken into a limited number of large plates moving relative to one another. Where the plates separate at the ocean ridges new lithosphere is created, which thickens with age and cooling. Where the plates converge—the island arcs and young mountain systems—old ocean lithosphere returns to the deep mantle. Where the plates slide against each other—as in California—major fault systems develop. Accompanying the interactions of the plate margins are earthquakes, volcanoes, and the concentration of mineral resources.

The importance of this model to the geological sciences can hardly be overstressed. It provided earth scientists for the first time with a working model of the Earth as a whole, a unifying concept of global structure and composition, a fresh context in which to view earth history, and a framework in which to set detailed local investigations.

In succeeding years, development of the model has led to very successful explanations for the development of the oceanic lithosphere and of the major topographic features of the oceans. Through comparative studies of the moon and the terrestrial planets, we have generated a picture of the early history of the Earth and why it appears to be the only one of these bodies on which plate tectonics processes have been active. We are less certain of the nature of the driving forces that move the plates, and we are only beginning to understand the manner in which the continents have been generated and assembled over the last 4 billion years.

In the last decade we have come to recognize that the continents are made up of a large number of microplates assembled over long periods of time. We have found that hydrothermal circulation penetrates deep into the crust and that on ocean ridges the return circulation to the surface may yield metallic sulphide deposits in remarkably short periods of time. We have discovered evidence of major singular events that may have affected the geological record and life on a global scale. We have become able to evaluate the heterogeneity of the mantle, both in physical properties and in chemistry, which has implications with regard to the driving mechanism of plate tectonics and to the formation of continents. In the framework of plate tectonics we have been able to construct generic models to assist us in our search for resources and for a better understanding of the causes of natural hazards.

Technological advances have accompanied these conceptual advances. We can now measure the properties of materials at pressures comparable to those at the center of the Earth. We can perform precise geochemical and isotopic analyses of very small samples of earth materials at levels of 10^{-9} to 10^{-10} grams. Techniques are becoming available that will allow us to monitor crustal motions and positions at the centimeter level of precision. Space techniques have been developed that have permitted us to map the gravity field over the oceans in surprising detail, revealing the fine structure of the sea floor. We can apply tomographic techniques to map the three-dimensional structure of the Earth's interior.

The diversity of approaches within the academic earth sciences is generally well served by the research grant programs of the National Science Foundation (NSF) and, to a lesser extent, of other federal agencies. The Panel strongly supports these programs. In addition, there now exist research opportunities, often large in scale, that are especially timely because technology has become available or because significant progress demands a new approach. Such programs, e.g., scientific drilling or establishment of seismic arrays, are costly and may require considerable investigator collaboration and coordination, but they offer the promise of major increases in our knowledge of the Earth, of how it works and has worked, and of its internal resources.

The scientific problems upon which the recommendations of the Briefing Panel are focused are as follows: (1) *Structure and Composition of the Lithosphere;* (2) *Dynamics of Tectonic Processes;* and (3) *Evolution of the Continental Lithosphere.*

The Panel has based its recommendations on relevant reports by committees of the NAS/NRC, especially the report *Opportunities for Research in the Geological Sciences* prepared at the request of the Director of the Division of Earth Sciences of the National Science Foundation (NSF).

Geoscience research activities of industry and of federal agencies are often closely intertwined with those of academia, although the goals may differ. Industry has developed technologies and data that are useful in enhancing our basic understanding of the Earth, and it has also applied ideas and techniques developed in academia to its own problems. Many university research activities supported by NSF are dependent on facilities supported or operated by other federal agencies; without these facilities the research cannot be done. At the same time, without support for this research, full advantage cannot be taken of the facilities and graduate students cannot be trained.

The Briefing Panel has identified five research areas in which significant dividends can be expected as a result of incremental federal investment in FY 1985. These five research areas are:

1. Seismic Investigations of the Continental Crust.
2. Continental Scientific Drilling.
3. Physics and Chemistry of Geological Materials.
4. Global Digital Seismic Array.
5. Satellite Geodesy.

ANALYSES OF RECOMMENDATIONS

SEISMIC INVESTIGATIONS OF THE CONTINENTAL CRUST

Seismic investigations include both reflection and refraction seismic surveys, which complement each other. Reflection probing of the continental crust has opened the third dimension to detailed view and has revised our concept of the deep structure of the continental crust and upper mantle. One technique is a modification of petroleum exploration methods. Because of the large service industry involved in the search for petroleum, availability of funds is essentially the only barrier to expansion of this type of exploration. At present only one field crew is being employed full time surveying for scientific purposes in the United States. An augmented effort would provide an opportunity to investigate some of the many attractive target areas that have already been identified as well as to experiment with new ways to apply the technique to obtain greater detail and increased penetration.

Seismic refraction and imaging investigations can use natural (earthquake) or artificial (explosive) sources to probe the continental

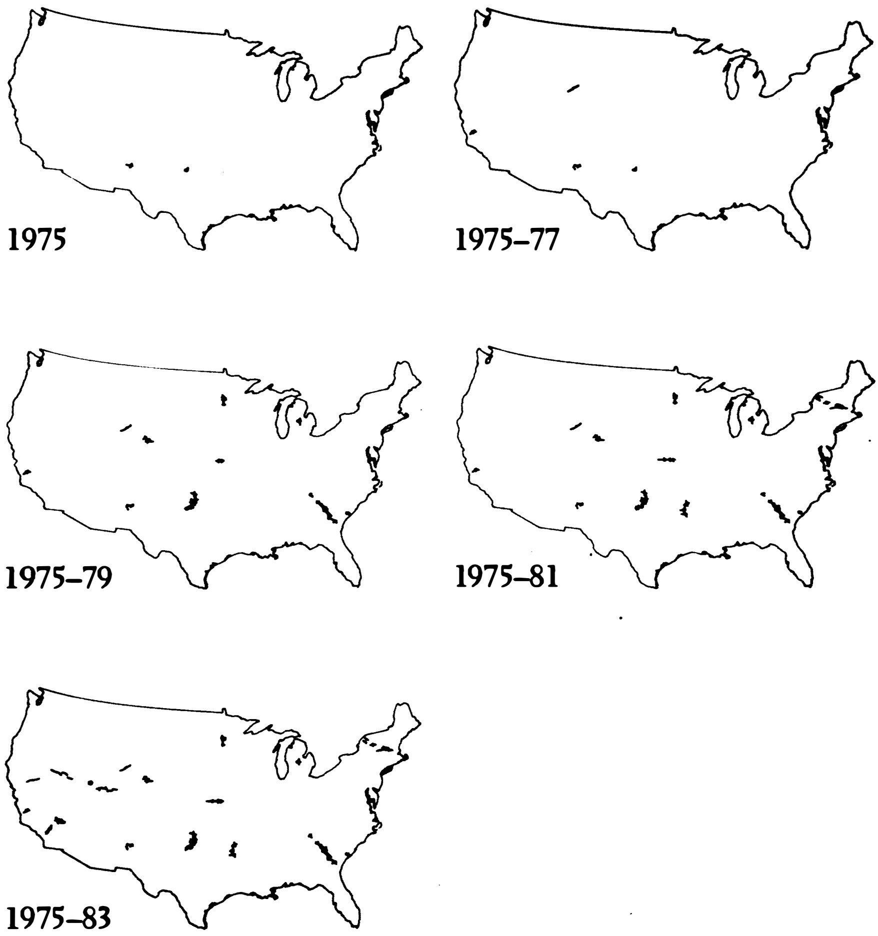

FIGURE 1
COCORP (Consortium for Continental Reflection Profiling) survey lines now total approximately 5200 kilometers.

lithosphere. In this case the research is presently equipment limited. Closely spaced arrays of up to 1000 seismic instruments are required to sample the lithosphere on a scale comparable to its geological heterogeneity. Operation of such an array requires considerable coordination and planning, and these have already begun. Capital investment in portable digital instruments should follow at a rate of 100 instruments per year, and when sufficient instruments have become available a large-scale experiment should begin. Some experiments using new three-dimensional imaging techniques will site the array for several months over particularly interesting areas for determination of deep structure.

The transition from continent to ocean is the fundamental discontinuity in the crust. It takes two major forms: passive margins, which are intraplate, seismically quiet transitions from continental to oceanic crust, and convergent margins, with earthquakes and volcanics, where oceanic lithosphere interacts with that of the continent. The U.S. Geological Survey (USGS), industry, and academic institutions have probed the upper few kilometers of both types of margins, but the deeper structure and the tectonic processes remain a mystery. Innovative applications of existing technology promise to delineate the details of the deeper structure and to allow us to reconstruct the structural, tectonic, and thermal history of these areas.

The most incisive experiments are provided by large seismic (reflection and refraction) arrays. These will require time on ships equipped with seismic reflection instrumentation comparable with that available to industry and arrays of ocean bottom seismometers similar to the continental array discussed earlier. On passive margins the continental crust extends as much as 200 kilometers offshore and often contains large sedimentary basins with resource potential. Seismic reflection surveying at sea will provide answers to critically important questions about the origin and evolution of these basins.

CONTINENTAL SCIENTIFIC DRILLING

Drilling in the oceans has revolutionized our ideas about the history and development of these areas; continental scientific drilling promises to be equally exciting.

In the report *Continental Scientific Drilling Program* (NAS/NRC, 1979), four principal problem areas for drilling are identified: thermal regimes; ore-forming processes; understanding earthquakes; and the composition, structure, and time and space relationships of the rocks of the continental crust. Panels of the Continental Scientific Drilling Committee (NAS) have identified priority targets in these four areas (see Figure 2).

The report *Opportunities for Research in the Geological Sciences* (NAS/NRC) recommends that NSF-funded dedicated continental drilling be initiated in support of numerous research opportunities. This is in addition to scientific drilling carried out by other agencies in support of their missions. Drilling is costly and must be based on well-founded plans for research before, during, and after drilling operations; thus the first requirement is for a national planning and preparation effort for holes dedicated to science. Several regional consortia have already been established, and efforts are under way to establish a national group of which these would be a part.

Many holes are drilled each year for purposes other than scientific research, and some of these are so planned and located that significant information could be gained from add-on experiments. Several such experiments have been carried out, but the absence of a dedicated budget for such experiments has made it difficult to respond to many of these opportunities in the short time frame that is usually available.

PHYSICS AND CHEMISTRY OF GEOLOGICAL MATERIALS

Measurement of the physical and chemical properties of rocks, minerals, magmas, and aqueous solutions under the extreme conditions of temperature and pressure found in the Earth's interior is required to understand the composition and flow in the mantle, the origin and evolution of magmas, and the evolution and circulation of hydrothermal systems. Similarly, the precise measurement of isotopic ratios of many elements, with techniques developed in the past few years, is needed to trace the secular evolution of igneous rocks and to establish the time constants of various geological processes. An extremely powerful set of techniques for identifying the major chemical reservoirs in the crust and upper mantle tracing materials through transfer processes and for estimating the age of formation of their derived products has been developed. These techniques utilize isotope chemistry and a variety of mass spectrometric instrumentation. A systematic survey of major global rock isotope and geochemistry systems would provide fundamental solutions to many major geological problems.

Progress in these areas depends not only on access to modern technologies and instrument development in the laboratories of chem-

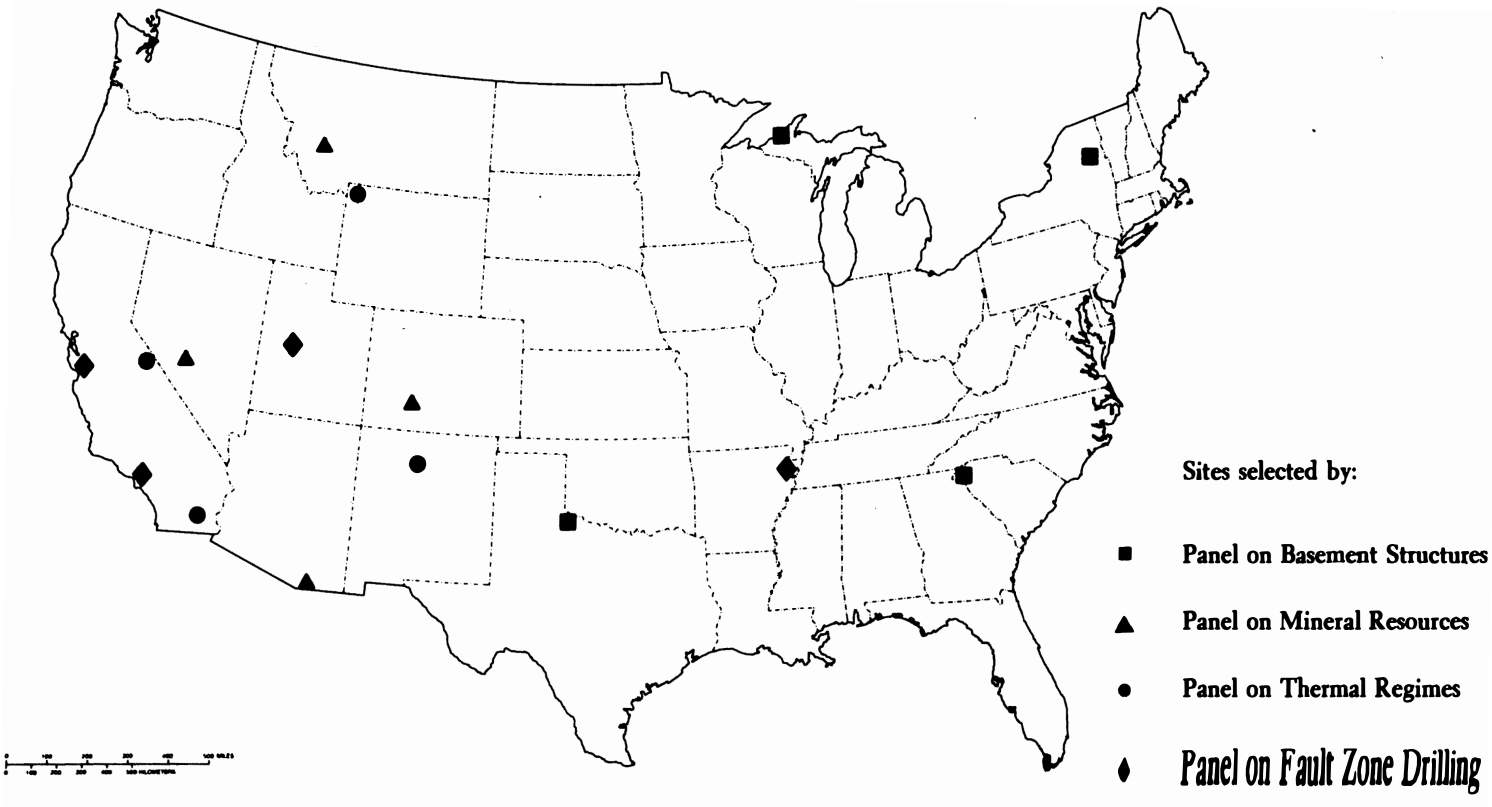

FIGURE 2
Proposed drill sites, continental scientific drilling committee.

istry, physics, and material sciences but also on the continuing development of these techniques to make laboratory measurements under the conditions of temperature, pressure, and natural contamination that characterize many geological environments. The high cost of instrumentation in this area and the limited funding available have inhibited the exploitation of rich opportunities.

This country has been a leader in developing the concept of plate tectonics, in applying innovative tools and techniques to studying the Earth and its neighboring planetary bodies, and in developing and manufacturing new instruments. In recent years this leadership posture has been eroded to some degree. A vigorous research program supported by the atomic energy and space efforts equipped many academic laboratories, but no comparable effort in instrumentation has been made for more than a decade. U.S. facilities now show significant obsolescence. If advantage is to be taken of the opportunities that are available in geochemical, isotopic, and high-pressure studies, a major commitment must be made to the development, purchase, and maintenance of new instruments and to the modernization of laboratories.

GLOBAL DIGITAL SEISMIC ARRAY

The seismograph array is the antenna by which the geophysicist receives signals propagated by the Earth. The global system of analog instruments has been the main source of data for seismological studies. A small number of digital installations are now in use, supported by DOD, USGS, and NSF; they have confirmed the enormous power of digital data for solving important problems. The goal is to determine the variation of velocity, anelasticity, anisotropy, and stress state through the Earth and, eventually, to map composition and convection.

This goal can be accomplished through a calibrated set of broadband instruments emplaced around the surface of the globe. An essential element is transmission (near real-time) of the data from about 100 stations by satellite telemetry to a properly equipped data center. All of the technology for achieving this network now exists, but significant support will be required to deploy the system, establish the central receiving and distribution center, upgrade a complementary analog facility, and fund the necessary supporting research.

The Panel recommends that the National Science Foundation act as overall coordinator and lead agency for funding such an array and that the operation be planned and overseen by a university consortium.

SATELLITE GEODESY

The new Global Positioning Satellite (GPS) system promises resolution of position to the centimeter level, giving us the capability of measuring actual plate motion and deformation rather than the results of these integrated over time. Use of GPS will revolutionize both conventional and tectonic, or time-dependent, geodesy.

To take advantage of this new tool, an investment will be required for construction of water-vapor radiometers and to purchase an initial set of GPS receivers for testing the accuracy against existing optical measuring instruments. In subsequent years support will be needed for the acquisition of additional instruments, for field observations, and for the operation of a satellite orbital determination network.

Satellite altimetry (measuring the elevation of the sea surface) recently has been used to construct maps of gravity anomalies in the oceans. These anomalies demonstrate, for example, that fracture zones are much more profound features than had previously been recognized, and they provide direct evidence for shallow convection in the mantle. These measurements will be continued by TOPEX, which is designed principally for oceanographic purposes. The resolution of these measurements can be improved by the systematic measurement of the gravity field from space, as proposed in the Geopotential Research Mission (GRM) of NASA. In addition, a mission such as this can extend high-resolution gravity (and magnetic) measurements to the continents. The value of this mission in providing global gravity coverage and better orbital data for TOPEX and other earlier missions makes it an attractive opportunity for an early start.

THE SCIENTIFIC OPPORTUNITIES THAT CAN BE EXPLOITED BY THESE RECOMMENDATIONS

STRUCTURE AND COMPOSITION OF THE CONTINENTAL LITHOSPHERE

In the last few hundred years the two-dimensional surface of the Earth has been explored both geographically and geologically. In this century the third dimension, depth, has been probed intensively in selected areas, particularly the oceans and sedimentary basins, but with limited resolution. Achievements with major impacts on science and society such as development of the concept of plate tectonics and the discovery of major deposits of mineral and energy resources followed this exploration. In this historical context, the next great frontier to be explored clearly seems to be the great body of rocks that form the continental crust and mantle.

Although many modern geological, geophysical, and geochemical studies are compatible with a comparatively simple layered crust, there are increasing indications of complex and discontinuous multilayers, some complexly deformed internally. Exposures of deep crust and samples of deep crust brought up by volcanoes indicate that a variety of sedimentary and volcanic rocks of surficial origin may make up a significant part of the lower as well as the upper crust.

In recent years it has become possible to apply the high-resolution geophysical techniques developed for petroleum exploration to the study of the entire thickness of the continental crust. Although, geographically, only a fraction of the crust has so far been examined, the results reveal a crust of far greater complexity than previously envisioned. In orogenic belts, thin slices of older rock have been thrust for hundreds of kilometers over younger sedimentary rocks of preexisting continental margins. This is an effect of the cycles of opening and closing of ocean basins that seems secondary in terms of basic plate movements but is of primary importance to the development of the continental crust. In areas of active rifting, such as the Basin and Range Province, steeply dipping faults forming and bounding the ranges are found to flatten and to merge at depths with shallow detachment faults that stretch for a hundred or more kilometers beneath the extensional terrain. Such faults may be reactivated zones of weakness from still earlier episodes of major compression and thrust faulting. What is perceived as the fundamental boundary between crust and mantle, the Mohorovicic discontinuity, is clearly far more complex and variable in both vertical and horizontal directions than was commonly thought, and it may vary widely in nature from one locality to another. Faults that may be conduits for mineral-bearing fluids have been traced from surface features, including areas of mineral concentration, to great depths. Molten bodies of rock (magmas) have been mapped at crustal depths where solid rock normally prevails. Deeply buried sediments that are worthy of exploration for petroleum have been found.

The continental crust does not disappear at the shoreline; it extends beneath the continental margins. The margins of the continents vary according to the nature of the processes that have shaped them. Rifting apart of the Atlantic Ocean has left behind a continental shelf in the United States where the evaluation for the potential of oil is dependent on better exploration by refraction and reflection seismic methods. The western margin of Central and North America contains the entire panoply of active plate tectonic processes; rifting in the Gulf of California, subduction of the Pacific plate beneath Central America and the Aleutians, and transcurrent motion from Mexico to northern California. Modern marine seismic methods have advanced rapidly, and these regions are now ripe for intensive exploration.

These new and striking observations of the crust revealed by seismic reflection profiling clearly indicate that a greatly enhanced understanding of the continents will be derived from further observations. Complementary information is derived from wide-angle reflection measurements, refraction measurements on the continents, and the application of newer imaging techniques. To provide measurements with adequate resolution, it is necessary to deploy portable instruments in very large, dense arrays, with up to 1000 instruments, much larger arrays than are currently available or that have been used in past investigations of the lithosphere. These arrays may be employed in refraction profiling using explosive sources, or they can be deployed to monitor earthquake activity over periods ranging from days to months. Such monitoring can provide locations of hypocenters so precise as to permit correlation with specific geological structures at a new level of precision—a level that would significantly advance the understanding of earthquakes and the tectonic processes that cause them. Distant seismic events recorded on these arrays provide the data for deeper imaging.

The ultimate exploration tool, as well recognized by the mineral and petroleum industries, is the drill. Deep-sea drilling has been invaluable in providing verification of geophysical interpretations and has provided detailed information on the structure, composition, and history of the ocean basins. A number of countries have established continental scientific drilling programs, including Canada, France, Belgium, Japan, West Germany, the United Kingdom, Czechoslovakia, and the USSR. The program in the USSR includes a scientific hole in the Kola Peninsula that is now at a depth of 13.5 kilometers (44,000 feet); drilling is continuing with a target of at least 15 kilometers. This is the deepest hole in the world (previous record was 33,000 feet in Oklahoma) and is located entirely in ancient Precambrian rocks. Most of the holes drilled in this country are drilled for water, oil, or minerals. Many are in anomalous areas, where resources are likely to be found; consequently, the basement rocks of the continental crust are unknown in large areas of the country. While some knowledge can be gained by attaching add-on experiments to the holes drilled for other purposes, it is becoming apparent that dedicated holes will be needed to fully understand the structure, composition, and age of the continental crust.

The continental crust and deeper parts of the lithosphere contain the record of about 4 billion years of earth history. The continental crust is also the repository of recoverable resources of minerals and fuels. Many segments of the present continental crust date from a time far back in geological history. There are good physical and chemical reasons to suppose that tectonic regimes and geodynamic mechanisms have changed with time. Important questions arise such as when and how rapidly the continental crust was generated, whether the process was episodic or continuous, and the rate of generation as a function of time.

Studies based on modern isotopic techniques suggest that we are at the threshold of being able to answer some of these questions. It now appears that the continental lithosphere is in some sense complementary to a residual mantle region of global scale. The rate and manner of extraction of continental crustal material from mantle sources can be understood by extensive isotopic and chemical studies of rocks from the upper and lower continental crust. These measurements provide the ages of and information on the nature and history of source materials from which the continental crust has been derived. Present isotopic data suggest that the formation of the continental crust is episodic and that very large volumes are formed in relatively short times.

Finally, in recent years it has been found that the differences between continents and oceans extend to considerable depths (150 kilometers or more) and that the deep structure of the mid-ocean ridges may extend to 400 kilometers. It is not clear whether these differences are statically or dynamically produced, but it is clear that understanding the nature of the continents will require exploration to great depths beneath them. Fortunately, new seismic techniques, similar to medical tomography, can provide this kind of information, on both a regional and a global scale.

DYNAMICS OF TECTONIC PROCESSES

Contemporary tectonic theories provide a framework for the synthesis of a wide variety of geological, geophysical, and geochemical data. However, these models are primarily descriptions of operative processes and are only slowly developing into rigorous and quantitative formulations. Increased amounts of new and high-quality data; advances in continuum mechanics, materials science, and other relevant physics; and availability of adequate computers make the present time right for rapid progress toward the development of more effective models of structural processes.

Research in tectonics is carried out in the field, where nature is observed; in the laboratory, where the myriad parameters in the real world can be controlled; and through theoretical studies. Laboratory studies of rock deformation tell us about processes in rocks and the changes that result from high deviatoric stress. The changes caused by stress provide a basis for interpreting geodetic and geophysical data in tectonically active areas. Seismic techniques are being applied to the determination of magnitude and orientation of stress in the mantle. Studies of petrofabrics in the field and in the laboratory provide clues to the temperature, stress, and flow conditions that prevailed during the formation of the rock. Thus we can study past as well as contemporary dynamics of the lithosphere and the underlying mantle.

The data critical to establishing the kinematics of plate motion were based on paleo-magnetic studies and marine geophysical measurements. Now, with the remarkable accuracy of new Global Positioning Satellite receivers, we can foresee that within a few years plate velocity vectors can be determined and accelerations in plate motion can be measured. The interaction of the plates in great earthquakes can be observed directly by measurement of earthquake displacements and post-seismic dissipation of strain. This new, relatively inexpensive technology provides one of the most exciting opportunities in the earth sciences today.

Knowledge of the stresses in the lithosphere is fundamental to understanding the physics of tectonic processes. Because it is a tensor quantity describing forces internal to a mass of material, stress is difficult to observe directly. Overcoring and hydrofracturing techniques have been used for in situ measurements of stress in rocks, but only the latter is useful at significant depths. The cost of hydrofracturing rises steeply with depth because of the need to drill a hole; therefore, stress measurements should be included in the observation program for all deep holes drilled for scientific purposes. New holographic techniques are being developed for down-hole applications.

Stress measurements at isolated points are the beginning step in analysis, but it is the spatial variation in stress that drives dynamic processes. Because earthquakes are the result of stresses at the source, both the radiated waves and the permanent deformation they produce contain important information on stress. Seismic source theory and wave propagation theory are sufficiently developed that the dynamic and geometric properties of the sources can be derived routinely from analysis of sets of seismograms. This is done in the time domain by matching the calculated and observed waves and in the frequency domain by interpreting the spectral characteristics. The greatest need is for high-quality digital data from earthquakes with a wide spread of geographical distribution and magnitudes. The technology exists; it is a matter of deploying the instruments and creating the data management facilities to serve the experimenters.

Digital data are also required for the analysis of elastic wave velocity anisotropy. Interpretation of anisotropy data is more ambiguous than waveform analysis because it can result from both deviatoric stresses and crystal alignment. Crystal alignment in rocks of the mantle is itself the result of flow; thus, it can shed light on internal dynamic processes. These anisotropy studies have the advantage that they can be used to probe any region through which the waves travel, not just tectonically active areas. Detection of anisotropy from wave polarization anomalies and surface wave dispersion is being recognized as a powerful tool for studying Earth dynamics; the availability of large quantities of broad-band digital data will enhance development of this approach. Similarly, the attenuation of seismic waves is controlled by temperature and dislocation structure and, thus, stress.

Finally, the response of rocks to stresses is governed by their rheological properties. More complete knowledge of the elastic and inelastic mechanical and thermodynamic properties of the materials in the crust and mantle is needed as a basis for modeling processes ranging from metamorphism to basin formation and mountain building. New laboratories need to be built to measure the elastic and anelastic properties of minerals at extreme conditions and to synthesize materials at high pressure and temperature. These are frontier areas of materials science.

Dynamics and Structure of the Earth's Interior

The Earth's mantle and core are the major components of its interior. From geophysical studies, experimental investigations reaching the temperature and pressure conditions of the core-mantle boundary, geochemical and isotopic analysis of mantle-derived rocks, and modeling, we have obtained a rough understanding of the composition and physical state of the mantle and the core.

Plate tectonics and related phenomena result from flow in the mantle. This flow is also the principal process by which heat escapes from the Earth's interior. Beneath a rigid outer shell—the lithosphere—the Earth's mantle, which responds like a solid to short-term driving forces, behaves as a fluid over geological time. Plate motions are the result of convection in the mantle, but the pattern of convection is controversial; some advocate whole-mantle convection, others suggest separate systems in the upper and lower mantle. It is not known whether mantle convection occurs on a scale that is small compared to the large-scale flow associated with plate motions and whether the time dependence associated with the plate configurations is the only time-dependent aspect of plate tectonics.

It has been difficult to place observational constraints on the flow in the Earth's interior. However, techniques are now available for mapping of the Earth's interior in three dimensions (elastic wave tomography). One approach involves the systematic inversion of large amounts of travel-time data of compressional and shear waves that have traveled through the Earth; the other involves modeling based on free oscillations and surface waves having periods sufficiently long to be affected by the deep interior. Both approaches require large computing facilities and a global array of broad-band digital seismometers. Higher-resolution studies require large transportable arrays.

Recent advances in theory and increases in computational capacity are available to help solve the fundamental flow problems. Most models of mantle flow have been based on a single layer behaving as a Newtonian fluid. Mantle silicates have strong temperature- and stress-dependent rheologies and the mantle is surely inhomogeneous in viscosity and probably in chemistry. Melting that occurs in the course of convection affects the buoyancy. It is now possible to compute flow in a layered system with these complications taken into account; preliminary results are providing insight into the dynamics of the interior. When continents are present on the surface it may be that steady-state never prevails. The required three-dimensional finite element calculations exceed the capacity of most university computers and computer budgets, but they must be made if the understanding of mantle dynamics is to proceed beyond the present qualitative stage.

In addition to seismological data, data on the global geoid, gravity and topography, crustal and sedimentary rock thickness, and heat flow are also pertinent to investigations of mantle flow, rheology, and evolution. Of special value among newly available data are satellite-borne radar measurements of the shape of the sea surface (geoid). There are satellite techniques for extending this high-resolution geoid to the continents. Systematic treatment of these data will make it possible to place bounds on the locations of density and velocity anomalies in the mantle. Explanation of the geoid is one of the most challenging problems of modern earth science. If we understood the geoid, we should know whether the rheology of the mantle is strongly temperature and stress dependent, whether the convection is layered or whole mantle, and whether convection is statistically steady. It has been suggested recently that the geoid has a memory—that its present shape reflects conditions in the distant past. If this is true, then we have the exciting prospect of being able to identify past configurations of the continents, ridges, and subduction zones.

Laboratory measurements of the rheology of melts and silicate minerals at high temperatures and controlled dislocation densities are an essential part of these studies. Experimental geophysics has gone through a decline in recent years, in part because of obsolescence of equipment. New laboratories, filling the gap between rock mechanics and ultrasonics, are required to provide the basis for interpreting seismic attenuation and anisotropy observations in terms of rheology, stress, and flow.

A major advance in our understanding of the mantle has been the discovery that the mantle is chemically heterogeneous on a variety of scales. Evidence of this is provided by the precise analysis of Pb, Sr, Nd, and He isotopes of young volcanic rocks from the oceans. These data show that the mantle sources that yield mid-ocean ridge basalts (MORB) have been isolated and depleted, relative to other sources of basaltic rocks, in their radioactive and other trace elements for a period on the order of 1.5 to 2.0 billion years. Furthermore, the mantle sources that produce MORB appear to be a distinct global entity: relatively homogeneous chemically but distinct from all other mantle sources that produce basalts. For example, basalts from major volcanic centers on oceanic islands and from continental interiors cover a broad geochemical and isotopic range. The sources of these so-called hotspot magmas are enigmatic. Investigations in this area have been limited. Few laboratories are working on these fundamental problems because of the expense of establishing and maintaining modern isotopic facilities and their supporting high-precision separation and analytic equipment. Modern labs have recently been set up in England, France, West Germany, Australia, and Japan, often by U.S.-trained investigators.

Magma Genesis and Fluid Flow

Igneous activity—melting in the Earth's interior, migration of magmas toward the Earth's surface, eruption on the surface or emplacement at depth—is the major process of crustal formation and one of the major mechanisms of heat and mass transfer in the outer regions of the Earth. It is the principal mechanism by which volatiles are transported to the surface to form the atmosphere and hydrosphere; cooling of magma bodies sets up hydrothermal systems that produce some of our most vital mineral deposits. Volcanic eruption of magmas may present a significant hazard, but the study of the solidified products of magmas provides clues about conditions in the Earth's interior and their evolution. Generation of magmas not only results from convection upwelling but also, through phase transformations, may influence the density structure of the mantle and provide a positive feedback on the convective process itself.

Investigation of magmas requires a strong combination of field and laboratory studies, but the most significant recent advances have come from development and application of tools from physics, chemistry, and material science to the problems. We can now, in the diamond cell, measure the properties of minerals and fluids under static pressures up to 1.5 megabars and from cryogenic temperatures to 3000° C; we can study the speciation and kinetics of aqueous solutions similar in composition to ore-forming hydrothermal solutions; we can study crystalline and amorphous materials from new perspectives using newly developed or accessible spectroscopic techniques and surface techniques from the materials sciences. These developments suggest that future advances will be led by the new types of and higher-quality data made available through application of innovative analytical and experimental techniques.

Theoretical studies of fluid flow have also played an important role in modifying current thinking about processes in the lithosphere. Fluid dynamic modeling of magma migration in the mantle and crust and of circulation of aqueous fluids in the crust, including the progressive chemical interactions taking place during fluid flow and their feedback on the flow processes, are in their infancy. Preliminary results suggest potential breakthroughs in understanding the factors responsible for mineral-deposit formation and many of the observed chemical and eruptive characteristics of magmas. Such models are critically constrained by geochemical and isotopic measurements on "fossilized" circulation systems and magmas and by measurements of permeability of rocks and the physical and chemical properties of rocks, magmas, and hydrothermal fluids. The systematic application of isotopic techniques to the study of natural stable and unstable isotopic systems is critical. Field investigations, including deep sampling by drilling, and laboratory studies of rocks and model systems are needed to provide realistic constraints and inputs into such studies.

High-Pressure Geophysics

Our understanding of the constitution of the Earth's upper mantle is based primarily on interpretations of seismic data and on the study of igneous and metamorphic rocks believed to have been brought up from relatively great depths by volcanic activities. At deeper levels (greater than 200 kilometers), where our only probes are seismic waves and interpretations of the geoid, realistic interpretations in terms of actual mineralogy, composition, temperature, and rheology require measurements of the physical properties of individual minerals and aggregates at temperatures and pressures relevant to the mantle. Elastic constants and their pressure and temperature derivatives, rheological properties, and phase relations are vital to these considerations. Facilities for synthesis of large volumes of high-pressure phases in order to determine thermodynamic and elastic properties are available (principally outside the United States), and new opportunities are now available for studying the physical and chemical properties in the diamond cell under extreme conditions of temperature and pressure. Such measurements include scattering and reemission (e.g., Raman, fluorescence), diffraction (e.g., X-ray, Brillouin), and absorption spectroscopy (e.g., Mossbauer effect, visible, and infrared). Such measurements are essential not only for interpretation of seismic and gravity data but also as input into realistic models of convection in the Earth's interior.

EVOLUTION OF THE CONTINENTAL LITHOSPHERE

Our concepts of the manner of evolution of the Earth and its life have been heavily based on the gradualist ideas of Lyell and Darwin. Recently this manner of biological evolution has been questioned, and it has been suggested that it is not gradual, but punctuated; singular events may play a major role in determining the evolution of organisms. When we consider the evolution of continents we find a similar situation. One can explain the shaping of the Earth's surface as the result of slow processes; mountains appear to be built in geologically short intervals of time separated by long periods of relative quiescence. The average integrated rate of plate movements, measured paleomagnetically, is a few centimeters per year, but current activity along the plate margins is episodic, with earthquakes followed by quiescence; periods of intense volcanism followed by periods of little activity. Singular events, such as the great explosions of Yellowstone, Long Valley, or the Valles calderas, located within the continental mass, are anomalous in terms of a simple gradualistic model.

The continents, with their great age and complexity, carry within them not only most of the Earth's exploitable energy and mineral resources but also clues to the Earth's long history. In contrast, although the rocks of the ocean's floor are relatively young (less than 200 million years), they provide us with the keys to plate tectonics. The great ocean rifts are the primary distillery for separation of the lighter fractions of the mantle into the ocean crust.

The ancient cores of the continents are fringed by younger belts that have been added in the last billion years. In some zones of collision, accretion to the continent of shingles of overthrusted oceanic crust and sediment has contributed to continental growth, while the denser mantle has descended below the continental lithosphere. Study of the accretionary process at the margins of the continents needs to be augmented by extensive examination through geological and geophysical methods and, ultimately, by drilling.

The steady accretion process is punctuated by collisions of continents and by episodes of rifting of the continental lithosphere to create new continental margins and new ocean basins. Sedimentary basins, often containing hydrocarbons, form over these rifts, and the sediments contain the evidence of the rifting process and subsequent subsidence. Rifting has been pervasive during the last billion years, but many of the rifts cease their activity and do not become continental margins. The eastern continental margin is an outstanding example of a successful rift. Its shallower portions have been studied extensively for scientific and petroleum exploration purposes, but understanding of its deeper regions and its full history await clarification by advanced seismic techniques and drilling.

The continental blocks now appear to have been assembled through the course of time from small fragments, some broken off other continents. Examples of this phenomenon have been best identified on both the eastern and the western parts of the country but are evident elsewhere. A broad spectrum of techniques can be brought to bear on identification of these fragments: paleomagnetism; geochemistry and isotope geology; field mapping; structural and stratigraphic studies; and, where the rocks are buried, seismic measurements and drilling.

Prior to one billion years ago, the evidence for modern plate tectonics processes becomes more obscure. This is not surprising, as the record itself becomes more limited. Processes other than those associated with plate tectonics as we understand them may have contributed to the formation of continental blocks. At the present, two opposing hypotheses can be entertained: that the processes of plate tectonics, operating over a long period of time, have caused the gradual emergence of continental crust from the mantle, and that most of the volume of continental crust was generated early in the Earth's history by processes lacking modern analogs.

To answer this fundamental question we will have to draw on all of our tools—geological, geophysical, and geochemical. Within the major parts of the crust, continent and ocean, and the mantle and core, large-scale geochemical heterogeneities or reservoirs have appeared since the formation of the Earth. The number, size, spatial arrangement, and evolution of reservoirs capable of yielding distinctly different magmas constitute a matter of vigorous debate. In addition, the nature and magnitude of fluxes between reservoirs are poorly known but are essential to understanding the chemical and differentiation history of the Earth. For example, some oceanic basalts have been interpreted as retaining geochemical signatures of recycled ancient ocean crust, raising the question of whether differentiation of the mantle is proceeding irreversibly or whether it is approaching a steady state with recycling.

Characterization of the reservoirs can be greatly advanced by the simultaneous measurement of the Nd/Sm and Rb/Sr isotopic systems together with measurements on the same sample of uranium, thorium, lead, and oxygen isotopes and trace-element abundances. A variety of other isotope systems is now being exploited. The measurement of rare gases, particularly $^{3}He/^{4}He$, has shown that mantle reservoirs may not be totally outgassed; components that have not previously been on the Earth's surface are still being supplied from the interior. The distribution and evolution of geochemical reservoirs are intimately related to the dynamics of the plate-mantle system and to the thermal and tectonic evolution of the Earth and its continents. The trace-element and isotopic chemistry of crustal and mantle magmas and rocks are essential petrological data, yet only a few laboratories in this country are equipped to make these measurements.

CONCLUSION

Early investigation of the Earth required keen observational powers and a knowledge of the composition of rocks and minerals, of the distribution of organisms through time, and of fundamental geological processes. Through these investigations we gained understanding of the near-surface geology and we were able to exploit this knowledge to obtain resources from the Earth. Similarly, we were able to determine the gross features of the Earth's interior, first through geophysical observations and, more recently, through geochemical and geophysical experimentation and isotopic studies.

The solid Earth interacts with all of the other domains of the geosphere: oceans, atmosphere, solar terrestrial. It is the sole source of information regarding the history of actions and reactions of the geosphere and biosphere. Improved knowledge of the solid Earth and the processes that operate within it are significant not only to basic science but also to society. We live on the solid Earth; we extract our resources from it; it produces the magnetic field through internal motions; from it came the gases of the atmosphere, the waters of the sea and the salts contained within them, and the nutrients that feed its organisms. Terra firma usually describes it well, but internal and external forces produce changes with significant impact on society.

Our tools have become sharper in recent years, and advances in theory and technology provide us with new opportunities to explore in greater detail the nature of and processes in the interior of the Earth as well as the development of the Earth through time. The availability of modern instrumentation and techniques will contribute to the training of the next generation of scientific leadership in the solid earth sciences and of the personnel sought by government and industry to assess hazards and resources.

RESEARCH BRIEFING PANEL ON COGNITIVE SCIENCE AND ARTIFICIAL INTELLIGENCE

William K. Estes *(Co-chair), Harvard University*

Allen Newell *(Co-chair), Carnegie-Mellon University*

John R. Anderson, *Carnegie-Mellon University*

John Seely Brown, *Xerox Palo Alto Research Center*

Edward A. Feigenbaum, *Stanford University*

James Greeno, *University of Pittsburgh*

Patrick J. Hayes, *University of Rochester*

Earl Hunt, *University of Washington, Seattle*

Stephen M. Kosslyn, *Harvard University*

Mitchell Marcus, *Bell Telephone Laboratories*

Shimon Ullman, *Massachusetts Institute of Technology*

Staff

Paul S. Rosenbloom, ***Rapporteur,*** *Carnegie-Mellon University*

David A. Goslin, ***Executive Director,*** *Commission on Behavioral and Social Sciences and Education*

Sarah M. Streuli, ***Administrative Secretary,*** *Commission on Behavioral and Social Sciences and Education*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON COGNITIVE SCIENCE AND ARTIFICIAL INTELLIGENCE

Start with a fundamental question that arose with the scientific world view: How can mind exist in a world totally governed by physical law? Shift that question, with the coming of experimental psychology: What laws and regularities govern human behavior? Convert that question, with the growing understanding of biology: How can the brain give rise to all the phenomena that we call mental? Expand that question, with the emergence of computer science: What is the general nature of intelligent action, independent of the device (brain, computer, etc.) by which it is implemented? All these questions reflect a single underlying great scientific mystery, on a par with understanding the evolution of the universe, the origin of life, or the nature of elementary particles.

Nothing short of a full history of Western intellectual endeavor can make sense of the crazy quilt of disciplines devoted to facets of this fundamental question, each in its own way. Cognitive science and artificial intelligence stand together in taking information processing as the central activity involved in intelligent behavior and in taking the framework of modern computer science as the foundation for understanding information-processing systems. Cognitive science reflects a concern with how humans process information, being an offshoot of cognitive psychology, linguistics, and philosophy. Artificial intelligence reflects a concern with how computers process information, being a part of computer science. The rapid growth of these disciplines in recent years is advancing our understanding of the nature of mind and the nature of intelligence on a scale that is proving revolutionary.

The research cited in this report reflects three main themes:

1. Answering the fundamental scientific question of the nature of intelligence and how it arises out of primitive cognitive functions. A subtheme, but one pregnant with hope, is the joining of this understanding with work in neuroscience, to make new progress in grounding the mechanisms of intelligence in the mechanisms of the brain.
2. Solving increasingly complex problems by constructing computer systems with the requisite intellectual, perceptual, and learning powers—making full use of the increasing speeds and memory being provided by hardware technology.
3. Developing human talent by combining an understanding of how humans learn and perform cognitive tasks with effective intelligent systems to aid this development.

THE DOMAIN OF COGNITIVE SCIENCE AND ARTIFICIAL INTELLIGENCE

How can we characterize this broad field? Look first at what investigators are typically found to be doing. An investigator in artificial intelligence is constructing a computer program that can detect trends and constancies in data and induce physical laws. The program (aptly called *Bacon*) has rediscovered Kepler's third law of planetary motion and Galileo's law for the pendulum, among others. The investigator is trying to produce a program that can accomplish a task requiring kinds of reasoning hitherto presumed to be the province solely of human intelligence. He is also progressing toward discovery programs of greater generality through the evaluation and improvement of the approximate rules (heuristics) used to guide the program in its search for physical laws.

A cognitive psychologist is measuring the speed with which people can carry out mental operations on visual images—for example, rotating a figure in imagination—and observes that the time required to rotate a figure mentally is proportional to the angle of rotation, as specified externally. The investigator is testing hypotheses about the form in which information is represented in images and working toward a model capable of simulating the way mental operations on images enter into reasoning and problem solving.

The scope of the combined field of cognitive science and artificial intelligence is much wider than suggested by these two illustrations. In one direction are formal investigations of the abstract properties of computational systems that are basic to all intelligent behavior. In another are experimental studies of human observers detecting signals of millisecond duration in noisy backgrounds. In yet another direction are analyses of verbal communications to find the rules whereby people resolve ambiguities in language.

The common core of these diverse enterprises may be taken to be the task of understanding intelligence and intelligent behavior. What might we expect of a general theory of intelligence? Most immediately, it should help us to understand how the constituents of intelligent function have evolved in the human mind, the limits they set on possible achievement, the potentialities that are open to improvement. At the same time, the theory should help us understand how computers can carry out human-like activities and the ways in which they may exceed human capabilities. More broadly, an adequate theory of intelligence should guide attacks on questions or issues that go beyond our ordinary experience. Considering the current searches by some physicists and astronomers for evidence of intelligent life elsewhere in the universe, would we recognize intelligence in organisms very different from ourselves? Is it possible to set any theoretical limits on the possible differences in level or quality of intelligent function? Is human intelligence subject to biological constraints that do not apply to artificial systems?

ORIGINS

Artificial intelligence has been a recognized field of study for only a little over 25 years. Its theoretical substrate was the development of ideas regarding mathematical logic and computation during the 1930s and 1940s. The rapid development of digital computers from 1950 on provided a challenge to get these machines to accomplish tasks that had previously been the province solely of human intelligence. The computer provided, for the first time, a mechanism whereby theories of intelligent performance could be rigorously defined, built, and tested —providing the feedback necessary for rapid progress. The first efforts were the well-known attempts to program computers to play intellectual games, notably chess. But as it became clear that intelligence was based on the ability to process symbols and that computers were general symbol manipulators rather than just high-speed numerical calculators, the scope of efforts rapidly broadened to include research into induction, problem solving, theorem proving in logic, language understanding, and vision.

Cognitive science emerged into public view in the late 1970s, as an interdisciplinary field to cover all the sciences adopting an information-processing approach.* It did not come out of thin air. Seventy-five years of work by experimental psychologists brought human cognition into the laboratory and developed models sufficient to enable meaningful measures of the speed of mental operations and the amounts of information stored in memory following a learning experience. However, the ability to address how people deal with problems much more complex than those of simple laboratory tasks came with two developments of the 1960s.

The first was a suitable conceptual framework for a science of cognition. Developments in computer science and artificial intelligence gave rise to the view of the human as an information processor—a special kind of processor perhaps, but still one sharing basic principles and constraints with computer-based information-processing systems. The computer made clear how to partition the staggering task of comprehending the human mind and brain in all their aspects into two more tractable components. The critical idea was the distinction between computer hardware and software. One can learn how a computer works by studying its hardware—its electrical circuitry, recording medium, and timing mechanisms. With little knowledge of such matters, one can master the software and become a highly skilled programmer.

In the study of human intelligence, the scientific objective is to construct a theoretical picture of the structures and processes of human cognitive function at the level corresponding to software—the theoretical notions referring to the way information becomes encoded and organized in memory, the cognitive operations available for accessing and transforming the information, and the functional principles governing the sequencing and combination of operations. Cognitive scientists continue to hope for the ultimate explanation of mental phenomena at lower levels of scientific analysis, but they have found

*As such, artificial intelligence is included in cognitive science; we have kept them separate here to emphasize the joining of human and engineering disciplines.

out how to proceed toward the development of useful theories without waiting on developments in neuroscience. These cognitive theories are thoroughly mechanistic, even though the mechanisms are not yet related to how they are realized in the brain.

The second main source of concepts and methods for the new cognitive science came from advances in linguistics. The importance of language in human thought has always been recognized in a general way. But the beginnings of a clear understanding of what it means to know a language, and therefore the first possibility of discovering how languages are learned and used, flowed from developments in theoretical linguistics in the 1960s, in particular Chomsky's formulation of generative grammars—sets of abstract rules that produce exactly the set of grammatical sentences. Increasingly sophisticated and psychologically relevant theories of grammar have provided a framework for interpreting studies of language-based cognitive activity and posed in clear and researchable form the major problem of how the understanding of language is achieved and how language is acquired by the young.

FROM THEORY TO RESEARCH

Research in artificial intelligence and cognitive science is strongly theory driven. Increasingly, empirically oriented research is aimed at finding evidence relative to general theories rather than being a process of accumulating empirical facts and attempting to induce laws or principles. Thus a look at theoretical issues concerning system design and function will lead directly to the principal lines of research within which we can identify special opportunities with a high yield from increased efforts.

SYSTEM DESIGN

In artificial intelligence, system design refers to the relation between the way computers are constructed and the kinds of programs that can be implemented on them. The key concept is the *architecture,* which is the structure of the machine that permits it to be programmed—thus, the structure that creates the distinction just discussed between hardware and software. Any particular part of a system can be implemented in either hardware or software, with the choice to be made on the basis of trade-offs involving implementation cost, speed of execution, frequency of use, functionality, and flexibility. A critical question is what mechanism must be included in the hardware in order for the system to be able to manifest intelligent capabilities. What mechanisms, if not built into the architecture, would make basic intelligent capabilities (such as learning and problem solving) too difficult or slow to be feasible?

We have come to understand that one major aspect of the architecture is that it provides for *symbolic* behavior. That is, it permits computers to have internal representations that *refer* to other things. The computer's symbols (which are basically its addresses and operation codes) seem special, referring only to data structures in its own memories and to its own information processes. Yet out of that mode of reference comes the ability to refer to things in the outside world —to objects, activities, events, and abstract concepts. The internal symbols access data structures plus programs that interpret the data structures as representing the external thing. These data structures and programs were constructed by processes that had information about the outside thing. Because of this we ourselves can program computers to deal with many things—anything to which we set our mind. But systems themselves can also construct such symbols and their meanings when given information about the external object. Thus, for the first time we see how symbols and symbolic behavior can be realized in a physical system.

The hypothesis that links artificial intelligence and human cognition together, providing the frame within which scientific inquiry proceeds, holds that all intelligence, including human intelligence, arises from the ability to use symbols. By symbols we mean the physical symbol systems, noted above, as first understood in computer science. Without this stipulation the hypothesis would constitute only another version of the long-held view that symbols are important in human mental life. But with the notions of symbol and symbol system firmly connected to the physical world via the realization in computers, the hypothesis anchors the study of all the higher processes of mind to the same scientific world of mechanism as all other natural sciences.

Much is known about the abstract structure of such symbol systems (so-called universal computational systems), so the linking hypothesis is not that somehow the human is like a computer but that both humans and computers embody these systems and gain, thereby, their ability for intelligence and information processing.

Whereas the architecture of a computer is determined by the designer, the architecture of the human cognitive system is given by nature, and discovering its properties presents a major scientific challenge. For instance, it is evident that a human does not carry out actions the way current computers do, by a simple interpretive cycle of fetching the next instruction and then executing it. Rather, the human is much more recognition driven, with the current situation (at each instant), including the goals and momentarily active memory, directly determining the next action to be taken. The human cognitive system seems to be highly general-purpose, taking on a particular shape only in the course of adapting to the demands of a specific task. But structural constraints—those aspects of the cognitive architecture that determine which intelligent functions are easy or difficult to perform—are clearly of major importance. We know the human brain has many highly developed special structures for specific functions. Resolving the nature of these constraints is a prime task for research in cognitive science.

SYSTEM FUNCTION

In artificial intelligence the computer is made to take on intelligent function by programming it within the framework provided by the architecture. The endless diversity of intelligent behavior stems from the ability of an intelligent system to program itself—to create symbolic structures to guide its own future behavior. Programmability per se does not make evident how the intelligence is attained, even given a suitable architecture. What sort of programming leads to intelligent action? This too is now clear in outline. The two basic ingredients are search and knowledge.

Search arises by defining a space of possibilities large enough to contain the sought-for solution and then searching for the solution by generating the states of the space. Thus arises the familiar exponentially expanding search of potential future chess positions by chess programs as well as the search of possible isomers of a given chemical formula by the *Dendral* program in searching for an isomer that satisfies the mass spectrogram data of an unknown sample, thus discovering its structural formula. Indeed, all the basic methods of intelligence appear to involve search.

Knowledge is necessary to guide the search through the space—to avoid having to find the proverbial needle in a haystack. A prominent component of artificial intelligence systems is the memory organizations that hold the knowledge and permit finding the right bit of knowledge at the right time. For instance, much of the progress in the development of expert systems—programs capable of exhibiting expert-level performance in intellectual domains—has come from discovering that large amounts of how-to knowledge can be effectively represented by collections of *if-then* rules (productions), where each rule of behavior carries the conditions in which it is to apply (the if-part) as well as what it is supposed to do (the then-part). In contrast to how-to knowledge, bodies of factual knowledge are organized into networks of associations, so that access to part of the knowledge base provides the connections to obtain other relevant parts. These networks go beyond the simple concept of association. For example, they must be organized so that widely applicable knowledge is not redundantly stored with each specific object to which it applies. Knowledge about an automobile must be composed of what is specific to the car (its dents) plus knowledge about all cars of that make. But the latter is similarly composed of knowledge specific to the make-plus knowledge about cars in general. This progression continues up through the most general nature of physical objects, imposing a hierarchical as well as an association structure on the knowledge. The discovery of these forms of representation and memory organization and their efficacy for wide classes of tasks constitutes important scientific knowledge about intelligent systems.

Beyond matters of organization is the question of content. For a computer system to solve problems about cars, it must know about cars —and not just the specific facts about Fords and Chevies but all the concepts that make up the domain, such as steering, carburetors, that motors heat up, what a motor is, what heat is, and so on. Artificial intelligence systems differ from more standard computer applications precisely in that the knowledge involved in the system is not highly circumscribed. By the same coin, our current artificial intelligence systems, despite much progress, are still woefully limited in their knowledge compared to what general intelligence requires. Obtaining the knowledge to incorporate in systems is itself a major scientific task. This is especially true of expertise. One of the interesting results to come out of the work on expert systems is that experts cannot explain in general to another person the knowledge behind their ability to make expert choices. Only when put into concrete choice situations can they evoke reasons for their particular choice, and these reasons must then be integrated into a total organization (as in the *if-then* rules above).

Intelligence is shaped to an important extent by the properties of the languages and representations that it employs. Human intelligence is intimately involved with natural language (e.g., English), and in this respect it contrasts with the programming-language representations available in computers. The extent of the resulting constraints on human thinking and the degree to which they can be incorporated in computers are not yet fully understood. Human thought is also involved to an important, and perhaps not fully appreciated, extent with imagery. A new line of research in cognitive science is the investigation of models that may enable us to understand the "language" of the imagery system.

MAJOR RESEARCH AREAS

With the background above, we can now turn to specific areas of research.

ARCHITECTURES

It should be apparent from the discussion so far that architecture is a key concept for cognitive science and artificial intelligence, bringing in its train a large number of significant notions. Research into architectures is an active and important part of the field.

Computer Architectures

The most active research on architectures is in artificial intelligence rather than cognitive science. This effort is concerned with discovering the best architectures to support powerful problem-solving and learning mechanisms. Not only must computers be created with immense power (artificial intelligence supercomputers), but many organizational issues in the architecture are still to be discovered and understood. For example, how should memory be structured and accessed so that contact can be made between a novel situation and stored knowledge and problem-solving methods? How can new information be integrated into existing memory structures so that the system can learn through experience? Research into such topics translates directly into more powerful applied systems. It also provides a fund of knowledge about possible mechanisms upon which theories of human cognition can be built.

Massive Parallelism

The study of massively parallel systems has recently emerged as a major focus of architectural research. Originally, of course, computers were primarily uniprocessors (and artificial intelligence research concentrated on aspects, such as problem solving, that were primarily serial). But for some time computer science has been devoting major attention to how to exploit parallel computation, driven by the increasing difficulty in developing ever faster circuits and attracted by the emergence of integrated-circuit technologies that favor parallelism. There are many flavors of parallelism, varying with the number of concurrent processors, the intimacy of communicative coupling between the processors (measured by the instructions performed between significant interactions), and the volume of data flowing between them.

Massively parallel architectures are designed to have many simple processors (millions), with each one dedicated to a conceptual entity represented in the system. This contrasts with current architectures (including parallel ones) that dedicate a segment of memory to each conceptual entity but share processing capacity by shipping each piece of data to a processor for its moment of modification. The immediate research problem is learning how to compute effectively with these million-fold parallel architectures. However, the research vista that structures of this new class open up for cognitive science and artificial intelligence is more than just the hope of increased power. This class of architectures makes new contact with biological computation, which is also massively parallel. Indeed, the work is being deliberately guided by biological systems, especially the visual system.

Human Cognitive Architecture

Whereas in computer science and artificial intelligence, research on architectures takes the form of designing architectures and investigating their properties, in cognitive science the approach must be to hypothesize possible architectures and discover whether their implications for cognitive processes accord with empirical observation. The task is formidable because the university of plausible hypotheses is immense, and properties of architectures are only remotely connected to observable behavior. A major step toward making the enterprise realistic was taken in the 1970s. The key notion is deriving models from cognitive theories that, when realized as running computer programs, not only embody the theory but also perform the cognitive functions. A (soundly conceived) model of how a human learns can itself learn, and a model of how a human solves problems can itself solve problems. This operational property provides an important link in the chain of inferences from theory to observation, allowing behavioral consequences of complex architectural mechanisms (and combinations of them) to be directly determined. Comprehensive models of the human cognitive architecture have not yet been attempted, but several useful models for substantial subdomains, such as verbal memory, reading, imagery, and skill acquisition, have been put together. This trend toward integration is fundamental to making basic progress.

SENSORY INFORMATION PROCESSING

Most of our knowledge of the external world comes to us through either our eyes or ears—vision and speech. Hence, understanding the workings of these systems is a prime scientific problem. It is also critical for technological reasons. Vision is necessary for autonomous robots or similar devices that could simulate human performance in industrial settings or in the exploration of space. Speech is necessary to provide the ultimate natural means of communication between humans and machines.

Visual Perception

The end product of the analysis of a visual scene is subjectively familiar. We recognize the objects present in the environment and their spatial relations to each other and are able to compute distances in a way that permits us to navigate; also, the visual experience leaves a residue that can somehow be revived in the form of visual images, which enter importantly into reasoning about natural phenomena. The very earliest stage of the process leading to these consequences is well understood. Light reflected from the visual scene activates a mosaic of receptor cells on the surface of the retina at the back of the eye. But how information about objects in three-dimensional space is extracted from the two-dimensional mosaic of stimulated points and transmitted through the nervous system over nerves that carry information only by virtue of rates of firing of nerve impulses presents a formidable problem. It is somewhat paradoxical that the retinal image contains an enormous amount of information (for example, an aerial photo is typically digitized into about 70 million bits) yet underdetermines the three-dimensional scene that gave rise to it. Thus a device that could build a representation of a scene in the manner of human vision must incorporate knowledge other than the momentary visual input. Discovering precisely what knowledge will suffice and what computation must be performed on the input to yield a three-dimensional representation is a prime task for cognitive science.

The early stages of visual processing produce representations of basic information, such as contours, texture, motion, and depth, that provide the basis for determining shape. All these pose very hard scientific problems, both to discover what the mammalian visual systems do and to produce computational systems with equivalent performance. On none of these aspects are we close to attaining human levels of performance. This is not for want of practical incentive; high-quality systems for any of these would be extremely useful in artificial vision systems and would feed directly into the development of robotics. Indeed, substantial artificial intelligence research has already occurred on them all.

The current lead problem of this kind is that of stereo vision, namely determining just how the brain combines the images of a visual scene coming from the two eyes and uses small differences between them to construct a representation of the scene in depth. However, each image contains a very large number of elements, and it is a most difficult problem to construct search algorithms to find the element of one image that corresponds to any given element of the other. Some progress has been achieved by what are termed relaxation methods, but stereo algorithms take inordinately long to run on available computers. Recent developments in computer hardware technology, using very large scale integrated circuits, now provide the opportunity for significant advance. Using very large numbers of parallel computations, the three-dimensional field can be generated efficiently from the two-dimensional mosaic of retinal receptor elements. There are high hopes that in the next few years this problem will be solved, and we will obtain computational performance in the realm of human performance. This would represent a signal achievement, opening the way to an assault on the other problems of this type.

Because the exploration of low-level vision algorithms is being driven by what we know about psychophysics and neuroscience, this is the area in which cognitive science and artificial intelligence are in closest touch with neuroscience and where the prospects for a highly profitable bridge are best.

Speech Recognition

Since the 1950s, recognizing speech has been a prime target both of speech scientists and of technologists trying to develop recognition machines that would enable easy communication between human and computer. Speech recognition might seem a simpler problem than vision because its input is only one dimensional, not two, but speech has additional complications stemming from the complex encoding produced by the vocal apparatus. What appears to the native listener as a discrete sequence of words is, viewed acoustically, a continuous signal without breaks. Each small sound interval is the result of several contiguous segments of the speech it purports to represent—posing a genuine decoding task. Much of the progress to date has come only with the drastic simplification to recognizing separated words rather than continuous speech—the current commercial market is built upon this reduction. The recognition of words in normal continuous speech remains a major open problem.

In speech, as in vision, ambiguity of perception can be resolved by using knowledge and search heuristics from the cognitive realm. Efforts to bring these resources to bear on continuous-speech recognition programs in the 1970s yielded substantial progress—but only to the point of handling limited vocabularies (1000 words) in systems with strong syntactic and semantic constraints. Recent empirical work, however, has demonstrated that human listeners can recognize continuous speech efficiently using only low-level knowledge of words and phonology, together with knowledge of how basic speech sounds are realized in the acoustic wave form. Implementing these results in artificial intelligence programs offers promise of a new burst of progress.

NATURAL LANGUAGE

The understanding of natural language has come to be one of the central problems of cognitive science. (By *natural language,* we mean language as learned and used in ordinary life, as distinguished from the formal languages of mathematics and computing.) One reason is the important role of language in human thought and communication, another the need for machine comprehension and generation of language in order to permit easy and effective communication of humans with computers.

Learning Natural Language

In dramatic contrast with the difficulty of constructing a computer program to understand unrestricted natural language is the speed and apparent ease with which young children learn to cope with language. Preschool children outperform any artificial intelligence program, and the average six-year-old builds new vocabulary at a rate of more than 500 words per month, at the same time perfecting, at an automatic and unconscious level, mastery of a complex of grammatical rules that enable comprehension and production of an infinite variety of sentences. Moreover, language comprehension requires a semantic analysis of the message conveyed by the language, and learning new concepts often parallels the acquisition of linguistic descriptions for these concepts.

Before we can expect to cope with the full complexity of this learning process, we must accomplish the more basic tasks of investigating the induction mechanisms and knowledge organization principles that give rise to human learning. Significant recent results on learnability, together with some new developments in linguistics having to do with lexically based grammars and their incorporation in models of syntactic processing, are beginning to yield an idea of some of the properties of a reasonable theory of language acquisition. However, the investigation of learning methods is an endeavor that transcends natural language to encompass the acquisition of new skills, concepts, inference rules, and knowledge representations.

Understanding Natural Language

Investigating how adult human beings (or machines) comprehend and produce natural language need not wait on the development of a theory of language learning—and, in fact, has progressed faster. Such research bears on the possibility of easy communication with machines via typed or spoken commands and on the way information is gained from text or discourse in education learning.

The study of natural language understanding can be viewed from three mutually supportive objectives:

1. The identification of linguistic universals and the codification of grammatical systems into principled theories providing a formal characterization of language. This objective is associated primarily with the field of linguistics but clearly affects the automatization of language comprehension by computers and the understanding of possible mechanisms that may underlie human communicative abilities.
2. The simulation of human linguistic performance with a view toward understanding human communication, the cognitive processes that support linguistic communication, and the inference system required to extract the intended meaning from the language. This approach is typically associated with cognitive science, and it affects the design and construction of practical language communication systems in major ways.
3. The development of computer systems that can interpret language in limited but increasingly sophisticated domains. This objective is the central artificial intelligence concern with natural

language—making computers able to communicate with humans untrained in the intricacies of any particular formal language.

Natural language understanding systems produce internal representations of the information conveyed by the system. The most useful internal representations are *canonical* (recognizing that there are many different ways of expressing the same thought in any language), *semantically rich* (encoding information inferred as well as information explicitly found in the surface form), and *unambiguous* (unlike the natural language itself). These international representations are then applied to a variety of tasks such as text comprehension, data-base query, command interfaces, story interpretation, etc.

Artificial intelligence systems have gradually expanded the capabilities for language comprehension by machine:

- *Lunar* was one of the first truly integrated systems that performed grammatical analysis and subsequent semantic interpretation to answer questions posed by geologists about rock samples obtained from the Apollo-11 mission. (For example, "What is the average concentration of aluminum in the high alkali samples?")
- *Lifer/Ladder* was the first semipractical system used in database query. The final version allowed some tolerance for user errors such as misspellings or the use of occasional sentence fragments.
- *Margie, Eli, Sam, Pam,* and *Politics* were based on a scheme of encoding semantic relations called conceptual dependency. Rather than striving to provide user interfaces, these systems attempted to emulate human comprehension and focused on performing inferences about the intents, plans, and possible outcomes of the speaker in a dialog or of characters in stories. This work demonstrated how language comprehension is an integral component of the total human cognition rather than a separable skill. The same mechanisms used for solving problems, planning actions, and recognizing situations come to play in interpreting and generating language.

Additionally, theoretical work in the structure of dialogs, such as speech-act theory, dialog focus, and belief spaces, has paralleled the implementation of working-language interpreters. Moreover, recent advances in linguistics, such as Lexical Functional Grammar, are unifying previous purely syntactic approaches (such as Chomsky's transformational grammar) with semantic interpretation methods. These parallel theoretical developments, together with recent work on robust parsing (i.e., understanding ungrammatical language) and advances in generating language from conceptual structures, promise to yield a new generation of more flexible and powerful language understanding systems. Such systems are starting to be prototyped as unified practical language interfaces to expert systems, data bases, operating systems, and other computational facilities.

Neither the artificial systems nor the models of human comprehension yet come close to a satisfactory treatment of full language understanding. Nonetheless, this area is one in which returns for investment in research are sharply on the upswing.

INTELLIGENT SYSTEMS

The recent highly publicized advances in artificial intelligence in developing so-called expert systems reflect progress in understanding how substantial knowledge can be used to perform tasks. They have used realistic tasks such as medical diagnosis, petroleum prospecting, and industrial assembly, and enough highly specific knowledge to almost entirely eliminate search. For example, *Mycin,* one of the classical expert systems, performs diagnosis for particular realms of bacterial infections (e.g., blood diseases) and does so at the level of a medical expert. *Mycin* is a collection of several hundred rules, each of which expresses a bit of diagnostic knowledge and each of which actively calls itself forth whenever it is relevant to advancing the diagnostic task. Each rule is applicable to only a few situations, but in the aggregate they provide a mosaic of coverage that produces, in effect, a field of responses that guides the program in all situations. This knowledge, extracted from expert physicians, represents codified know-how. Though strands of medical science weave through it, just as in all clinical medicine, it is really the accumulated wisdom-applied-to-action of the experts, not their ability to reason or understand.

Systems with Full Expertise

The progress into programming with lots of knowledge, rather than with lots of search (or with highly repetitive algorithms, as in scientific computation), although it has already opened up interesting domains of application, simply sets the stage for the next round of important research. One immediate issue, for instance, is how to weld together extensive knowledge and extensive search in a single system in which the spaces of possibilities to be searched can be created and evoked wherever appropriate.

Another major issue is how to provide such knowledge-intensive programs with a deeper understanding of the substantive domain of their expertise, so they can reason their way to conclusions and can explain why its advice is sound. (Present systems do provide some explanation, but it is relatively shallow, in accord with the superficial character of their knowledge; for instance, *Mycin* can present the rules used in a particular diagnosis to explain what it did.) A substantial body of active research on the nature of such basic understanding is concerned with examining how human experts understand their domains of expertise. The cognitive science research into the nature of expertise is in close contact with the more applied work on expert systems being pursued within artificial intelligence. As a result, we expect the current rather shallow expert systems to metamorphose into much more capable systems—and more scientifically satisfying examples of intelligent agents.

Learning

Studying simultaneously how intelligent systems work and how they acquire their capabilities is so large a task that over their early decades both artificial intelligence and cognitive science resorted to fractionation. In each case the first focus of effort was on the static aspects of information processing.

Artificial intelligence has so far made its major gains by concentrating on performance—on what is needed by way of representation, however, and search to accomplish tasks of varying degrees of difficulty and specification. Nonetheless, the problem of learning is a bottleneck for both practical and theoretical reasons. Expert systems with millions of knowledge rules can be generated only by learning, and in truly intelligent systems learning must go on concomitantly with all problem solving. So until learning is understood, we are walking on one leg. Sparked by these needs, and a better grasp of the theoretical and computational tools required to approach the problem, a resurgence of work on learning is under way.

In cognitive science we have just come through a period in which attention has been focused almost exclusively on the end products of learning—the way acquired information is represented and organized in memory. With some significant results in hand, we are seeing a better balance, with the emergence of studies in concept learning and skill learning. The former line is elucidating how people learn both formal and natural categories (for example, logical or mathematical as compared to biological categories or colors). In the latter line, new experimental approaches in conjunction with models are yielding insight into the way extremely high levels of skill are attained in some areas (for example, the performance of mnemonists and "lightning calculators").

The increasingly close intertwining of the artificial intelligence and cognitive science approaches raises prospects of significant advances in our understanding of learning in all of its various guises. The results can be expected to spill over to the full field of artificial intelligence, as learning mechanisms are incorporated into performance programs in myriad ways, and should amplify the applicability of cognitive science to problems of education.

SCIENCE AND MATHEMATICS EDUCATION

Educating people in technical subjects is a good example of how work from many parts of cognitive science and artificial intelligence comes together. It is both a currently active area of research that shows great promise for making fundamental advances and an area that is capable of making solid contributions to the nation's education problems. Current distress with early science and mathematics education is due in part to social and economic factors but also in part to aspects of learning and instruction where contributions might be sought from cognitive research and theory. The issues and potential contributions are somewhat different in two major problem areas.

Tutorial Systems

One of the roots of difficulties at higher levels of education is generally believed to lie in our continuing inability to produce acceptable skill at reading, spelling, elementary arithmetic, and the like in the full population of normal school children. A similar problem is becoming acute in the training or retraining of adults in the basic knowledge and skills needed for employment in jobs with increasingly high technical demands.

From the standpoint of cognitive theory, what might be a source of our persisting lack of success at producing basic skills in the full population of normal school children? A firm generalization arising from research on learning of skills is that acquiring significant competence in domains of any difficulty requires large amounts of guided practice—much more than is provided for most pupils in schools or most adults who try to prepare for new technical jobs.

The combination of advances in cognitive theory and basic advances in computer technology (providing interactive workstations of substantial power at increasingly modest costs) makes possible significant inroads into the problems posed by technical education. What can be expected of computer-implemented instruction at this juncture? An intelligent tutoring system that can provide genuine help in educating a student in some well-understood domain, such as mathematics or science, must provide several components.

1. A powerful model of the task domain, so it can itself solve problems in that domain.
2. A detailed model of the student's current level of competence, encompassing both partial and erroneous competence as well as perfect competence.
3. Principles for interpreting the student's behavior, so as to be able to infer the student's knowledge and difficulties.
4. Principles for interacting with the student, so as to lead the student to a higher level of competence.

Experimental tutoring systems already exist for significant aspects of geometry, algebra, and arithmetic, and others are currently under development for some aspects of computer programming and elementary physics. For instance, a system dubbed *Euclid* works with a student who is learning to prove theorems in geometry—querying the student, evaluating progress, and providing advice.

The work in tutoring also feeds back and contributes to basic understanding of cognitive science and artificial intelligence. Recall *Mycin,* the artificial intelligence expert system mentioned earlier. The *Mycin* analysis of its task domain of diagnosis made it a candidate for an automated tutor capable of instructing medical students. An attempt to construct such an intelligent tutor showed that *Mycin*'s knowledge was too purely diagnostic know-how and that *Mycin*'s problem-solving style was too different from appropriate human strategies. A revision of the *Mycin* system (called *Neomycin*) attempts to provide a much more cognitively relevant treatment of the task, thus leading back to relevant research on the expert system side.

As artificial intelligence provides expert systems in an increasingly wide variety of tasks, and as the models of students' performance and learning proliferate, the potential exists in this area for a substantial contribution to education in science, mathematics, and technology. Additional elements not yet technically feasible, such as natural language interaction, should soon become possible. Progress on natural language interfaces for tutoring systems is expected to move at a more rapid pace than general research on language comprehension because the constrained nature of the domains to be taught allows simpler, though less general, techniques to be used.

Mental Models and the Problem of Transfer

To have an adequate supply of college-age students ready to move toward careers in science or engineering, effective education in mathematics and science must have begun much earlier. But it is a hard fact that these subjects are regarded as difficult, and very large numbers of children, perhaps including some of the most talented, are turned off by their early experiences. Furthermore, results of cognitive research suggest that faulty starts may be very costly. Inappropriate modes of thinking instilled at an early age, or during early experience with new material at any age, can be very difficult to modify later.

When solving problems in their field, experts mentally construct *qualitative* models of their tasks and reason within these models by applying knowledge and search. For example, when trying to solve elementary physics problems, expert physicists reason qualitatively about the situation until it becomes clear what calculations to make. They construct a diagram of the physical situation and reason to the variables that determine the motion. In contrast, naive problem solvers attempt to solve such problems by going directly to reasoning mechanically with the equations.

The ability to do qualitative reasoning goes hand in hand with expertise, but qualitative models per se do not necessarily lead to successful problem solving. The physicist's qualitative models are useful because they are built around the concepts of force and energy. When naive problem solvers try to bring their own expertise from daily life to bear on physics problems, their qualitative models look more like the earlier pre-Newtonian impetus models (a ball shot from a circular pipe continues to curve, although gradually straightening out). These naive theories provide a framework within which novices interpret demonstrations intended to teach them new facts and begin to learn to solve quantitative problems. Unless these informal theories are understood by the teacher and painstakingly supplanted by new habits of qualitative reasoning about natural phenomena, they persist as a barrier to learning. Thus, even after conventional instruction in new physical, or other scientific, concepts, students who do well on textbook examples often prove incapable of applying the laws or formulas they have mastered at a formal level to interpret actual physical events. The proper sequencing of experiences with qualitative and quantitative reasoning is, consequently, a critical aspect of early science and mathematics education.

The difficulty of producing transfer of learning from educational experiences to new problems has been a bugaboo since the earliest days of educational psychology. The first major research results discredited the prevailing idea of "formal discipline," the notion that general memory and problem-solving abilities could be strengthened by such measures as instruction in geometry and the classics. But nothing was left in its place except demonstrations of the extreme specificity of most learning to its original context. (In World War II, prospective aerial gunners were given intensive practice on skeet ranges—the sole discernible result being skill at breaking clay pigeons.)

It has become clear that for materials or tasks of any complexity, certainly including all forms of science and mathematics learning, a key to transfer is conceptual understanding. Important recent advances in cognitive science include empirical methods for assessing a person's qualitative understanding of principles in a domain (e.g., motion of objects on surfaces in kinematics) and characterization of the mental models, akin in many respects to scientific models, that individuals form when learning about causal relations in such domains. One intensive new line of research attacks the problem of just how the mental models and modes of problem representation of the truly expert solver of problems in specific areas of science or engineering differ from those of the less competent. Beyond the immediate objective of attaining theories of competence comparable to those now available in linguistics is the longer-term goal of tracing the developmental path from the mental model of the novice to that of the expert. A by-product of this effort is a principled basis for adaptive competency testing—on-line interpretation of a student's problem-solving efforts that yields an assessment of understanding. The methodology is advancing rapidly and seems capable of extrapolation from simple mathematics to such complex practical tasks as nuclear plant operation.

NEEDS AND OPPORTUNITIES

Realizing the emerging opportunities for important theoretical and practical advances depends on the degree to which some important needs for support are met. Currently, much more support is available to computer science and artificial intelligence than to cognitive science.

This imbalance stems directly from their perceived roles as being hard and soft sciences, respectively. Yet it should be clear from this entire report that the two disciplines closely interlock and contribute to each other at every level and on every scientific problem. Cognitive science especially needs resources to permit it to contribute as a full partner.

The situation in basic artificial intelligence research has a different character. Surrounded by the sudden expansion of applied artificial intelligence, with its rash of new venture-capitalized firms and industrial laboratories, it is hard for basic artificial intelligence research to compete for the talented scientists. Even much basic research is seen by those doing it as very close to real application; hence, bright young scientists are tempted to move toward the commercial sector. From a broad social perspective this somewhat frenetic activity is merely the concomitant of obtaining the social payoff from science. But it tends to deplete the next generation of basic science. The only solution we know is the standard one of attempting to make the basic research situation sufficiently attractive to counter the dispersive factors.

Two specific factors have tended to dampen progress and may do so more seriously in the future. One is a shortage of computing power, and the other is a dwindling supply of new talent. These affect both cognitive science and artificial intelligence, although from somewhat different sides, as noted above.

Computers are the laboratories of artificial intelligence research and major components of laboratories in cognitive science. Unfortunately, both new theoretical ideas and new possibilities of application are outrunning the computing power needed to implement them. Many well-trained cognitive psychologists and other behavioral scientists are unable to move into newly active lines of research in cognitive science because they do not have access to the computing power and languages required for the development of theoretically significant systems. Even in the research groups where the most intensive efforts in artificial intelligence are going on, what appear to be significant opportunities for new theoretical advances are having to wait until some way is found to give investigators access to substantially increased computing power.

Two distinct kinds of computing power are needed. One is the networked powerful personal computer capable of substantial list processing. Adequate research environments in artificial intelligence now require essentially one machine per active investigator (which includes a substantial fraction of graduate students). The other kind is special hardware systems or artificial intelligence supercomputers (not to be confused with numerical supercomputers, such as the Cray and Cyber machines). These systems are all experimental at the present time and can be built only in computer engineering environments and often only in cooperation with industry.

A factor that is not yet critical but that might seriously constrict the advance of research within a few years is an insufficient supply of new talent. Theoretically significant research in cognitive science and artificial intelligence requires extensive technical training in mathematics, psychology, and computer science. Young people are unlikely to go through the rigors of this training unless there are prospects for research careers. The dwindling opportunities for faculty appointments in cognitive science at research universities is by no means compensated by the increasing opportunities in industry, for the balance of support for applied versus basic research is quite different in the two cases. If we want many talented young people to embark on the road to expertise in cognitive science, it would need to become known that a commitment to the support of basic research in this area on the part of the government and universities is sufficient to provide reasonable prospects of research careers.

The needs of this field for carefully engineered support are well balanced by the emerging opportunities for significant theoretical and practical advances. A number of promising possibilities are apparent in the research areas surveyed in this report.

1. The development of comprehensive and executable architectures for cognition.
2. The understanding of massively parallel computation, both at the level of computer architecture and at the level of models for cognitive processes.
3. The development of effective programs for computer recognition of visual scenes, language, and speech, drawing on the resources of both artificial intelligence and cognitive science.
4. The research that attempts to bridge the gap between both artificial intelligence and cognitive science and neuroscience.
5. The application of what is being learned about high-level human reasoning and processes of skill and knowledge acquisition to the development of a new level of expert systems in artificial intelligence.
6. The development of computer-based intelligence tutoring systems, which can play an important role in educating and reeducating our people for a high-technology society.

However, we do not wish to place too much weight on specific lines of research. Cognitive science and artificial intelligence differ from various predecessors most importantly in the emphasis on general problems of fundamental theoretical significance. Though a full theory of intelligence is still remote, the insight that both intelligent machines and their human counterparts can be seen as instances of the class of symbol-manipulating systems has yielded some general results and opened the way for more.

Broadly viewed, research in both cognitive science and artificial intelligence is showing that the acquisition and organization of information are basic aspects of intelligence in both humans and machines. A salient and pervasive characteristic of human beings, distinguishing them sharply from lower organisms and special-purpose computers, is a tendency to acquire information that goes far beyond current task demands and to organize the accumulating knowledge so that it can be accessed and used in unforeseeable future problem situations. Building this feature effectively into computer systems can be seen as the next major advance in the power of intelligent systems. Cultivating it during human cognitive development may be a key to materially increasing human problem-solving abilities.

An important offshoot of research on artificial intelligence is its clarification of the problems and its development of methods pertaining to the systematization and codification of the enormous accumulation of knowledge in technical domains. It is coming to be widely believed that information processing will be the heart of the next industrial revolution. (Witness the scale of the Japanese Fifth Generation Computer project, centered on very large scale machines for artificial intelligence.) As occurred earlier for energy and the physical sciences, and later for medicine and the biological sciences, only massive basic research efforts can prepare us to keep up with this fast-moving worldwide development.

But machine intelligence alone will not suffice. Human intelligence must comprehend the emerging problems and the machines that aid in their solution. Thus, new challenges face our already heavily burdened education system. Cognitive science may be able to provide the broadened theoretical framework within which significant progress can occur and artificial intelligence major components of the needed new educational technology.

RESEARCH BRIEFING PANEL ON IMMUNOLOGY

Hugh O'Neill McDevitt ***(Chairman),*** *Stanford University School of Medicine*

Barry R. Bloom, *Albert Einstein College of Medicine*

John J. Farrar, *Hoffmann-La Roche, Inc.*

C. Garrison Fathman, *Stanford University School of Medicine*

Mark I. Greene, *Harvard Medical School*

Leroy E. Hood, *California Institute of Technology*

Kimishige Ishizaka, *The Johns Hopkins University School of Medicine*

Richard Lerner, *Scripps Clinic and Research Foundation, La Jolla, California*

Philippa Marrack, *National Jewish Hospital and Research Center, Denver, Colorado*

Stuart F. Schlossman, *Dana-Farber Cancer Institute, Boston, Massachusetts*

Jonathan W. Uhr, *The University of 'Texas Health Science Center at Dallas*

Irving Weissman, *Stanford University School of Medicine*

Staff

C. Garrison Fathan, ***Rapporteur,*** *Stanford University School of Medicine*

Barbara Filner, ***Associate Director,*** *Division of Health Sciences Policy, Institute of Medicine*

Linda Peugh, ***Administrative Secretary***

Talitha A. Shipp, ***Secretary***

Gail Wilson, ***Secretary .***

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON IMMUNOLOGY

INTRODUCTION

Four questions have long intrigued immunologists:

1. How does the immune system have the ability (or the potential ability) to construct an antibody to virtually any foreign material —thousands and thousands of distinct foreign molecules, some not even found in nature but constructed in the laboratory?
2. How does the immune system recognize "self" and accept it while rejecting that which is foreign?
3. How does the immune system's memory work (the ability to respond more quickly and more strongly when challenged with a foreign material it has "seen" previously)?
4. How do the several cell types involved in the immune system communicate so that an integrated, regulated response effectively removes foreign material?

In the past two decades, immunologists have gained substantial understanding about the biochemical structure of the antibody molecule and the processes that enable the genetic system to generate such a great diversity of antibodies. They have learned of two major cell types—B cells and T cells—and have inquired deeply into the regulation of functions of the B cell. The T cell has been more elusive. However, recent advances in basic knowledge and some powerful new technologies have created major new scientific opportunities to identify the role of the T cells involved in regulation of immune responses and to understand the surface molecules (receptors) and secreted molecules (lymphocyte hormones) that mediate necessary interactions.

This new knowledge is leading to such a sufficiently clear understanding of the workings of the immune system that it can now be manipulated toward quite precise experimental and clinical goals. We can study the molecular mechanisms not only of the immune response but also of other basic biological processes, particularly those entailing cell interactions mediated by surface molecules or by minute amounts of protein signals.

Such recently acquired knowledge and techniques also have tremendous potential for the improvement of human health, for making possible effective new means of diagnosis, therapy, and disease prevention. In fact, the potential already is moving toward realization at the level of clinical experimentation.

The technological tools that have allowed these opportunities to develop include monoclonal antibodies, cloning of genes, cloning of immunocompetent cells, and automated and highly sensitive sequence analysis and synthesis of nucleic acids and peptides.

The major research opportunities for the near term are the T cell receptor and lymphocyte hormones (lymphokines), both activators and suppressors.

This paper will explain immune responses, the power of the technological tools, and the reasons why the T cell receptor and lymphocyte hormones would be significant and fruitful areas for scientific inquiry. The paper concludes with a discussion of the additional opportunities that would be offered by this new knowledge—opportunities to advance basic knowledge of the immune system, its responses and regulation of those responses; opportunities to advance knowledge of other basic biological processes and of disease processes; and opportunities to exploit this knowledge with new approaches to the prevention, diagnosis, and treatment of an array of diseases, such as autoimmune disorders, cancer, genetic disorders of blood, infectious diseases, and virtually any health problem for which organ transplantation offers a solution.

IMMUNE RESPONSES

When the body is challenged by a foreign substance (antigen) that is extracellular, such as bacteria, certain cells of the immune system respond with proteins (antibodies) that recognize and bind to the antigen. That response, called "humoral immunity," is mediated by B cells, which secrete a large number of antibodies into the blood. The antibodies mobilize many other immune system activities to rid the body of the foreign substance. A different kind of immune response is at work in intracellular infections, allergic reactions, or rejection of transplanted organs. This entails "cell-mediated immunity," a function of T cells.

Both B and T cells are lymphocytes originating in bone marrow. Each B or T cell recognizes a specific antigen and can synthesize a unique receptor that recognizes the antigen. Upon appropriate contact with the antigen, the B or T cells are stimulated to divide. They produce a "clone" of identical cells; cells in the clone produce identical receptor molecules and thus have identical specificity for antigen.

One of the great research successes of the past two decades has been the answer to the first question we posed: How does the immune system construct unique antibodies against a seemingly infinite array of foreign substances? The determination of the amino acid sequences of the protein chains in antibodies* and parallel genetic studies revealed that part of the immune cell's DNA is rearranged in order to provide the requisite antibody diversity. In other words, genetic endowment is not rigid and the precise structure for each antibody molecule is not inherited. Gene parts are provided, but they are sorted and resorted to form many new combinations.

This surprising discovery opened the door to a truly revolutionary view of all genes—one that encompasses control of expression during growth and development by rearrangement of a portion of DNA from one site to another along the length of a chromosome. Recent research has shown that this is indeed true for a number of genes, not only those determining antibody structure.

*A typical antibody molecule consists of two pairs of protein chains—two identical heavy (long) chains and two identical light (short) chains. Each chain, whether heavy or light, has what are called "constant" regions and "variable" regions. The constant regions are the same for a large number of antibody molecules, even though their specificity varies. This part of an antibody molecule confers properties held in common, such as helping to trigger inflammatory responses. The variable region, which provides specificity for antigens, is what makes each antibody clone unique.

MAJOR TECHNOLOGICAL GAINS

Recombinant DNA technology and relatively stable and homogeneous B cell lines were essential for the advances in knowledge of B cells. With new technology, the time has now arrived to investigate T cells, their receptors, and their regulation. The relevant technological achievements are the production of monoclonal antibodies, the cloning of genes, and the cloning of immunocompetent cells. Advances in sequence analysis and chemical synthesis of nucleic acids and proteins offer additional promise for the future.

MONOCLONAL ANTIBODIES

In 1975, Kohler and Milstein showed that certain malignant mouse cells that produce antibody molecules (myeloma cells) could be fused with normal antibody-producing B cells. The resultant "hybridoma" cells can be grown continuously and indefinitely as cell cultures that synthesize large amounts of antibody molecules. The antibody produced by such a clone of hybridoma cells is called "monoclonal antibody" and has the important property of having a single specificity—all of the antibody molecules are identical.

Because of unlimited possibilities in producing hybridomas, there is the potential to produce antibodies against a seemingly endless variety of important antigens. There is the technical hurdle of finding a clone with the desired specificity for antigen, but considerable ingenuity has been exercised in screening and selecting for clones of interest.

The discovery of hybridomas and monoclonal antibodies launched a new era in immunologic research and in the application of immunologic assays to basic and clinical problems. One research application has been in the identification of T cell types and elucidation of their biological functions. Other research applications cross a broad spectrum of biological areas.

CLONING OF GENES

The genetic information specifying the structure of protein molecules, such as enzymes and antibodies, is encoded in a cell's DNA. It is translated into a working copy—messenger RNA—for use by the protein synthesizing apparatus of the cell. Advances in genetic engineering (recombinant DNA technology) have enabled scientists to isolate messenger RNA from cloned cell populations with defined characteristics, such as the synthesis of a particular protein, make a DNA copy of the genetic blueprint inherent in the RNA, and then incorporate that DNA copy into the DNA of a virus. Monoclonal antibodies can then be used to detect those viruses that have "expressed" the protein of interest during their infection and reproduction in appropriate cells. Once the virus and then its DNA are isolated, the gene of interest can be purified.

Thus, by using recombinant DNA technology and monoclonal antibodies, one can isolate in very few steps the genes that encode proteins of biological and clinical interest. Antibody-directed gene cloning should lead to the rapid isolation of almost any gene, provided the researcher can define a functional polypeptide (protein-like molecule) and the cell that produces it. These immunologic and recombinant DNA technologies should continue to produce a rich diversity of reagents for identification and isolation of biological macromolecules of scientific, clinical, and commercial interest.

In particular, gene cloning should enable immunologists to study in detail the function of genes expressed selectively in particular cells of the immune system. Included would be antigen receptors located on the surface of T cells, other receptor systems that govern interactions between lymphocytes and other cells, and the hormones that regulate the growth and differentiation of the immune system's cells.

CLONING OF IMMUNOCOMPETENT CELLS

Recent discoveries have made possible three different approaches to the long-term growth of T cell clones in cell culture. One method adapts the hybridoma technique first developed for B cells and fuses malignant T cells with normal T cells. A second approach depends on the addition of "growth factors" to the cell culture medium. A third approach involves repeated stimulation of the cultured T cells with a particular antigen and serial reculturing.

These three approaches have set the stage for the identification and isolation of a variety of cell types and secreted cell products involved in regulatory events mediated by immunocompetent T cells. Preliminary experiments using similar approaches suggest that it will be possible to readily clone other sets of immunocompetent cells, including B cells and antibody-presenting cells. This suggests the eventual possibility of constructing an integrated in vitro network of the cells that govern immune responses.

CHEMICALLY SYNTHESIZED ANTIGENS

Proteins, which consist of long chains of amino acids folded into a three-dimensional structure, trigger immune responses and thus are active as antigens. Certain peptides, which are short chains of amino acids, also can act as antigens. Some chemically synthesized peptides have been found to mimic the relevant structure in natural antigens and to elicit immune responses.

What once took the efforts of several laboratories over a period of years (e.g., determination of the amino acid sequence of a single protein chain) can now be done overnight or in a period of days with automated machinery. The technologies that enable rapid and accurate analysis of the amino acid sequence of proteins and the chemical synthesis of peptides offer opportunities to probe the basis for antigenicity. In more applied terms, these technologies also offer opportunities to produce exceedingly safe new vaccines that completely lack an infectious agent. Chemical synthesis also provides the possibility of tailoring an antigen to be a more potent immunologic stimulant. New vaccines against hepatitis, polio, rabies, and herpes viruses are being designed using this approach.

For the future, there is no reason to restrict research to peptide mimics of antigens. Organic chemicals should be considered as well. If we learn the rules by which proteins can be mimicked by organic compounds, we can think in terms of entirely new classes of drugs. This would have applications not only in vaccine development but also in the treatment of hormone and enzyme disorders.

RECEPTORS AND SIGNALS THAT REGULATE THE IMMUNE SYSTEM

The immune system has two main attributes: recognition of self and nonself and immunologic memory. Both recognition and memory are a function of the receptors on the surfaces of immunologically competent lymphocytes and of the protein products elaborated by such cells.

The receptors on B cells that enable them to recognize and respond to foreign materials have been known for some time. They are immunoglobulin (antibody) molecules, closely related to the antibody molecules these cells secrete after stimulation. The protein structure of these molecules and the genes encoding them have been among the most rewarding areas of immunologic research in recent years.

THE T CELL RECEPTOR

Characterization of the antigen-specific receptors on T cells has been elusive. There are two major reasons for the difficulty: (1) T cells, unlike B cells, do not secrete soluble receptor materials, and (2) the receptors on T cells bind only to antigen when it is associated with proteins on the cell surface membrane, not to free antigen. This second phenomenon, identified by pioneering work in the mid-1970s, has been termed "restricted recognition." The cell membrane proteins involved in T cell recognition are the products of a collection of genes that constitute a locus called the major histocompatibility complex (MHC). Restricted recognition of antigen by the T cell suggests that its responses will be directed only at cell-surface-associated antigens. A cytotoxic T cell, for example, will kill a virus-infected cell that bears virus antigens and MHC products on its surface but will not interact with circulating free virus. (Free virus will be removed by antibody that is the cell-free product of B cells.)

For the immunologist the property of restricted recognition makes it difficult to describe the nature of the antigen-MHC protein complex recognized by T cell receptors. Indeed, it is still not known whether the T cell has two different types of receptors, one for antigen and one for MHC, or a single receptor that recognizes some sort of aggregate of the two molecules. The phenomenon of restricted recognition means that binding to antigen or to MHC cannot be used as an enrichment technique for the isolation of T cell receptors for a particular antigen. The situation has prevented our understanding how one-half of the immune system (T cells) recognizes antigen and regulates the function of the other half (B cells).

Two of the new techniques outlined above, however, have contributed to the recent isolation of T cell receptors for antigen-plus MHC by several laboratories. One technique depends on B cell hybridomas; the other employs cloned T cell lines or T cell hybridomas with receptors of a single specificity for antigen-plus-MHC. Several laboratories have shown that mice immunized with cloned T cell lines or hybridomas produce antibodies that either block or mimic recognition of antigen-plus-MHC by these T cell lines. In every case published to date, the antibodies are specific for the immunizing clone and do not interfere with antigen-plus-MHC recognition by other T cell clones, even those with similar specificities. Apparently it is the variable portion of the T cell receptor that is being recognized by the relevant antibodies.

The discovery of the T cell receptor for antigen-plus-MHC is so recent—and so much remains to be done—that it is difficult to predict exactly what advances will be realized. Certainly it will lead to the isolation and understanding of the genes encoding these products and to knowledge of the structure of the antigen-plus-MHC product that the T cell receptor recognizes. Additionally, it is known that certain cell surface molecules are associated with T cell receptors for antigen-plus-MHC and contribute to the overall interaction between T cells and their targets. These molecules—LFA-1, OKT4, and OKT8—may have genetic variants that affect their function. Isolation of these molecules and receptors will enable us to understand their function and perhaps to control their action.

Basic understanding of some disease processes also can be foreseen as possible extensions of this research. Many inherited diseases in humans are genetically linked to the MHC, including some so-called autoimmune diseases, such as juvenile-onset diabetes, rheumatoid arthritis, and multiple sclerosis. Because recognition of antigens by T cells is associated with recognition of MHC products, it is thought that diseases of this type are caused by some sort of abnormal T cell response to antigen-plus-MHC, which leads to cross-reaction with a self-antigen-plus-MHC. Understanding how T cell receptors recognize antigen-plus-MHC should lead to clarification of autoimmune phenomena and perhaps to development of methods of circumventing them.

Additional potential therapeutic applications can be envisioned. If the structure of the T cell receptor is known, and monoclonal antibodies can be raised against unique determinants of it, it should be possible to delete selected clones of T cells. For example, such techniques could eliminate clones of leukemic T cells or cells involved in organ graft rejection. Such an approach has the advantage of leaving the rest of a patient's immune system intact.

LYMPHOCYTE HORMONES

The recent discovery of a new class of molecules elaborated by lymphocytes has led to new developments in understanding cell regulation in the control of immunity. Under the proper stimulus, lymphocytes can produce molecules that stimulate other cell types to divide and replicate. These molecules are called "interleukins," which are a type of lymphocyte hormone. One interleukin active in tumor cells has been defined at the gene level, has been molecularly cloned, and has been shown to promote lymphocyte division. This linkage of immunology to cellular and molecular biology should provide valuable insight into the regulation of cell growth and cell division.

With interleukins, immunologists have been able to undertake cellular cloning of T cells. The clones of T cells grown under the influence of these molecules can now give insight into the molecular organization of the recognition system of these uniquely specific cells. Furthermore, it has been found that clones of lymphocytes obtained from mice exposed to tumors and grown under the influence of interleukin-2 can retard and eradicate a similar tumor in other mice. It is anticipated that such interleukins (of which three distinct types have already been defined) will have application to tumor therapies. Interleukin-3 has been shown to promote the division of certain blood cells, suggesting that their growth may be regulated in part by this type of molecule. It is anticipated that interleukin-3 will facilitate the study of blood production by the body and lead to therapeutic modalities for some blood diseases.

Another area of study relates to the expected discovery of other interleukins whose growth-producing activities are restricted to other cells of discrete lineages. Although discovery of interleukins capable of stimulating activities of lymphocytes and other blood cells is clearly a major advance, it also may be possible to identify new genes or to molecularly alter already defined interleukin genes so as to make their hormone products capable of stimulating preferentially one type of cell whose absence needs to be corrected. Furthermore, these interleukins may help provide insight into signals and activities important in the growth of normal and malignant cells. They also may lead to a new family of pharmaceuticals with broad biological activities.

In contrast to the activities that promote cell growth and activity, another class of lymphocyte hormones that suppress cellular activities has been discovered. One type of lymphocyte, termed suppressor T cells, produces molecules that inhibit the development and expression of immunity. These molecules, which still require extensive biochemical definition, are thought to be polypeptides that operate in a sequence of cellular interactions to inhibit both B cell and T cell immune responses.

Areas of study for suppressor hormones relate primarily to the ways in which suppressor cells become activated by antigens, to the biochemistry of the molecule produced, and to the action of such cellular products in the inhibition of cellular processes. Previous research has been hampered by the very low concentrations of some of the suppressor hormones—insufficient amounts of hormone were available for biochemical and biological analysis. But the detailed study of the structure of suppressor lymphocyte hormones will undoubtedly be facilitated by B and T cell hybridoma technology. If both hybridomas that secrete homogeneous products and monoclonal antibodies for these products are developed, it should be possible to clarify the structure of the hormones. It also will be possible to determine the genes that encode these inhibitory substances and to study their organization.

The further development of human suppressor hybridomas is important. Since allergic diseases are caused by antibodies in a class called IgE, selective suppression of IgE antibody formation would be the fundamental treatment for these diseases. Human T cell hybridomas that produce IgE suppressor factors or that could be used for cloning of genes controlling such factors would provide a new strategy to control allergic diseases. Human hybridomas capable of elaborating nonspecific suppressor molecules have recently been defined. Thus, the development of more specific human suppressor hybridomas may have therapeutic application to transplantation and inflammation in addition to allergic responses.

Some mouse hybridomas secrete specific substances that inhibit inflammatory responses similar to those that occur in many pathological states of human beings. The development of human suppressor substances may prove useful in the treatment of autoimmune, allergic, and connective tissue disorders, in which abnormal humoral and cellular responses contribute to morbidity.

Suppressor cells may inhibit myeloma growth, suggesting that lymphoid dysplasias may benefit from new types of suppressive therapy. However, in other solid tumor models, suppressor cells can become activated and inhibit eradication of tumors. Thus, it is likely that another area of development will include strategies to bypass or regulate suppressor activity, which could lead to therapies for human cancers.

APPLICATIONS OF NEW IMMUNOLOGIC KNOWLEDGE

These advances in knowledge of the immune system will play a crucial role in furthering our understanding of basic biological processes and will have medical applications as well. The basic biological processes studied certainly will include the mechanisms and control of immune responses but also will include virtually any biological process involving cell-surface alterations, cell-surface interactions, or cellular communication via peptides present in very low concentrations. We have indicated already the additional important applications to basic research involving isolation of genes.

BASIC RESEARCH

Over the next decade there will be a rapid accumulation of antibodies reactive with many cell types and cell products. Already there are monoclonal antibodies that distinguish the two major classes of lymphocytes in the immune system, T and B cells, and that recognize subsets of these cells specialized to perform different functions. Antibodies will be developed to react with the new lymphocyte hormones and their receptors on the target cells. The ability to distinguish the multiple cell types and cell products participating in immune responses will help to elucidate the systems that control these responses.

It is likely that monoclonal antibodies will be generated that react with cells of many organ systems and even with subsets of cells in those organ systems. These monoclonal antibodies to various cell-surface antigens would constitute a library of molecules that distinguish cells in a precise way that was hitherto impossible. It is expected that the surface antigens will help identify different phases of function, different stages of maturation, or different lineages of differentiation. This identification will help elucidate the processes involved in control of diverse biological processes.

MEDICAL APPLICATIONS

The potential diagnostic, therapeutic, and preventive applications of this new knowledge in immunology are too numerous to discuss here in detail. New approaches are apparent for autoimmune diseases, cancer, and genetic defects in the blood-forming system, for example. Management of infectious diseases, especially via vaccines, and virtually any disorder necessitating organ transplantation also will benefit.

Examples of the readily foreseeable medical uses of monoclonal antibodies, lymphocyte hormones, and recombinant DNA technology are provided by juvenile-onset (insulin-dependent) diabetes mellitus and cancer.

Juvenile-Onset Diabetes

This disease affects 0.5 percent of all children in the United States. It leads to lifelong dependency on injected insulin and to a number of severe long-term complications. Recent research indicates that juvenile-onset diabetes is due to an autoimmune response—antibodies against their own insulin-producing cells are made and the cells are destroyed. It is possible to detect an autoimmune response months or even years before actual development of clinical diabetes.

We already know that most children with this disorder have two particular transplantation antigen types, DR3 and DR4. However, screening is not yet feasible because fully 35 percent of the normal population has these two antigen types. Further biochemical and molecular genetic analysis should identify subtypes that would enable prospective identification of a much smaller population at risk. Screening would then be feasible.

Isolation of the target molecules on the insulin-producing cells would open the way to specifically suppressing this autoimmune response. Use of this target molecule or of human monoclonal antibodies directed against the transplantation antigens involved in the development of autoimmunity might permit therapy that suppresses the autoimmune responses.

Thus, it should be possible within a few years to identify in advance a relatively small percentage of the population at risk for this disorder. These children would be tested on a regular basis for development of the autoimmune response, and treatment to prevent the development of diabetes could be initiated after early signs of autoimmunity.

In the case of juvenile-onset diabetes, monoclonal antibodies seem to offer the best opportunity for identifying, treating, and understanding this autoimmune disorder. For other autoimmune diseases in which there is too much or too little of a lymphocyte hormone, hormone analogs or purified hormones offer possibilities of restoring normal immune regulation.

Cancer

Monoclonal antibodies that are completely pure, have a single specificity, and show high affinity should aid in the discovery of tumor-associated antigens. Antibodies to these antigens offer opportunities for diagnosing tumors that are not precisely identifiable by conventional techniques and could even become the basis for new cancer taxonomies. These antibodies could be used to stain frozen sections, to stain cells dissociated from biopsy material, for tumor imaging, and to analyze blood using flow cytometry. Monoclonal antibodies to tumor-associated antigens also could be used to develop very sensitive radioimmunoassays that will detect these antigens in various body fluids.

One of the major problems will be to determine the distribution of any tumor-associated antigens on normal tissues. If the antigens are not unique to a given tumor, it may nonetheless be possible to use a panel of monoclonal antibodies tailored to assay for particular cancers. The results should allow earlier and more accurate diagnoses.

Monoclonal antibodies also offer new approaches to cancer treatment. They could be used for direct antitumor activity or for targeting delivery of toxic chemicals to tumor cells. For example, there are many highly toxic proteins produced by bacteria or plants—the plant product ricin, for example. A single molecule is sufficient to inactivate protein synthesis and to cause death of the cell in which it is present. Experimental studies indicate that toxins coupled to a tumor-reactive antibody are highly useful in ridding bone marrow of tumor cells. (In the autologous bone marrow rescue approach to the treatment of cancer, the patient's own "cleansed" marrow is transplanted back after massive doses of antineoplastic drugs or irradiation to kill tumors.)

The importance of bone marrow transplantation must be emphasized. In addition to therapy for certain cancers, it is the most obvious means for gene replacement therapy. In genetic deficiency diseases, especially those involving blood cell abnormalities, specific genes could be transferred into bone marrow stem cells, which would then reconstitute the normal functions of the recipient. But even when the donor and host are matched at the major histocompatibility complex, graft versus host disease has a rather high mortality, and this has impeded significantly the use of bone marrow transplantation. Fortunately, animal experiments and preliminary clinical trials indicate that antibody-conjugated toxins directed against T cells are capable of "cleansing" bone marrow of those cells that are responsible for the graft versus host disease.

Infectious Diseases

We close this discussion of medical applications with a brief mention of infectious diseases. Recent developments make it possible to purify antigens to an unprecedented extent for use as vaccines. Several methods are feasible, including using monoclonal antibodies to help isolate and purify antigens from natural sources. (One procedure often produces a 5000-fold enrichment in a single step.) Monoclonal antibodies and recombinant DNA technology also allow the genetic engineering of vaccines, and chemical synthesis is possible as well. Yet another approach entails antibodies made against the active site of a specific antibody. Occasionally the anti-antibody mimics the antigenic structure that elicited the specific antibody and engenders an immune response.

An example in which these techniques have been spectacularly effective is the production of a vaccine against the infectious form of the malaria parasite. The sporozoite can be obtained only from the salivary gland of mosquitoes, which do not lend themselves to vaccine production. However, a protective antigen was identified by a monoclonal antibody, the genes of the parasite were cloned in *E. coli,* and recombinant clones expressing the malaria-specific protective antigens were prepared. The vaccine has been found to protect experimental animals.

GROWTH OF IMMUNOLOGY RESEARCH

The truly impressive accomplishments in this field were made possible by the previous national investment of research resources. Vigorous growth in the past five years can be seen in a number of measures. For example, between 1977 and 1981 investigator-initiated research grant applications (RO1) increased more rapidly in immunology than in other fields of biomedical research without compromising quality. (There was a 27 percent increase in the number of competing proposals reviewed by the four immunology study sections of the National Institutes of Health but only an 18 percent increase in other fields.) Furthermore, between 1977 and 1982 the number of new Ph.D.s in immunology increased steadily each year, rising from 101 to 150. We believe that with adequate research resources the scientific opportunities presented in this briefing paper will sustain continued vigorous growth.

RESEARCH BRIEFING PANEL ON COMPUTERS IN DESIGN AND MANUFACTURING

Michael Wozny ***(Chairman)***, *Center for Interactive Computer Graphics, Rensselaer Polytechnic Institute*

Thomas Binford, *Computer Science Department, Stanford University*

Robert Brown, *Electro-Optical and Data Systems Group, Hughes Aircraft Company*

David Grossman, *Automation Research, IBM Corporation*

Christoph W. Klomp, *Computing Systems, Boeing Commercial Airplane Company*

James F. Lardner, *Deere and Company*

Michel A. Melkanoff, *School of Engineering and Applied Science, University of California, Los Angeles*

Stuart Miller, *Automation Systems Laboratory, General Electric Company*

Frank Stoneking, *Engineering and Manufacturing Computer Coordination, General Motors Corporation*

Arthur R. Thomson, *School of Engineering, Cleveland State University*

Herbert Voelcker, *College of Engineering and Applied Science, The University of Rochester*

Staff

George H. Kuper, ***Executive Director,*** *Manufacturing Studies Board*

Peter Smeallie, ***Staff Officer***

Delphine D. Glaze, ***Administrative Secretary***

Frances M. Shaw, ***Secretary***

Courtland Lewis, ***Rapporteur,*** *Herndon, Virginia*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON COMPUTERS IN DESIGN AND MANUFACTURING

SUMMARY

America's future growth in industrial productivity depends vitally on the health of the nation's research in computers for design and manufacturing. In spite of the impressive apparent growth, the panel notes that the central issue in this field is a pervasive lack of scientific knowledge. Universities have made major contributions in the past 10 years with very limited resources.

The first priority of this field is to build a genuine classical research community through increased government support of basic research. The panel identified the following five key research areas that will significantly impact the field:

- geometric modeling and analysis
- human-computer interface for design and manufacturing
- knowledge-based or expert technology
- information management
- manufacturing and computer devices

The second priority is to build our university educational programs in an effort to strengthen America's technical excellence for the long term.

INTRODUCTION

Although the use of computers in design and manufacturing is relatively new, the rapid growth of computer technology and the promise of improved productivity and quality have drastically changed the way we can design and manufacture products. Clearly, issues associated with the application of computers in the factory workplace transcend technology and impact industrial organization, management, and business strategy. Computers will have a profound impact on the work force at all levels, including managers, professionals, and production workers. This report, however, is limited to technical research opportunities for increasing the use of computers that must be explored if the United States is to maintain a competitive edge in design and manufacturing.

RESEARCH GOAL

At the risk of oversimplifying, we can think of design and manufacturing together with marketing as three inseparable parts of a production cycle from conception to delivery. Marketing provides information to engineering and design on the product required. Engineering designs the product and transfers product-definition data to manufacturing. The planning function within manufacturing transforms the product-definition data, e.g., the geometry, into process-definition data—the instructions for building the product—and then transfers this information to the factory floor. Computer-based planning and control systems directly schedule and track the manufacturing process as well as assure controlling the product's quality with regard to its design specifications.

Research on computers in design and manufacturing should lead to an understanding of the principles necessary to automate completely the creation, analysis, transmission, and management of all product-definition, process-definition, and related business data. This will ultimately take the form of a computer-integrated system. A corollary of this goal is to automate the physical processes corresponding to those data.

BACKGROUND

University research has contributed significantly to the advancement of this field. For example, the modern concept of computer-aided manufacturing (CAM) had its start in 1949–50 at the Massachusetts Institute of Technology under an Air Force contract. Because of the need to machine aircraft components with complex geometries accurately and reliably, the project led directly to the development of numerically controlled (NC) machine tools. An NC machine uses numeric or symbolic instructions to perform a manufacturing function such as milling or grinding. A subsequent separate effort developed the Automatically Programmed Tool language, called APT, a computer language still used for programming NC machine tools.

During the early 1970s, university research efforts in the U.S. and several other countries created the basis for solid geometric modeling, a complete representation of an object that permits any well-defined geometric property to be calculated automatically. This concept removed a major roadblock to flexible automation and has tremendous near-term potential to revolutionize the industry. The U.S. university research was funded primarily by the National Science Foundation.

FOCUS ON MECHANICAL TECHNOLOGIES

There exist two distinct technology areas where computers have been used effectively in design and manufacturing. One technology encompasses microelectronic chip design and fabrication. The other technology is associated with mechanical parts such as machines. Both technologies are very important. Although many of the generic manufacturing concepts, such as planning and scheduling, presented in this report apply to both technologies, the specific research needs for microelectronic design and process modeling are very different from the primarily geometric needs for mechanical design.

This report concentrates on the mechanical technologies because they represent an extremely broad range of diverse manufacturing applications whose mathematical models are not well understood. Mechanical technologies offer significant research opportunities, the results of which will benefit countless small industries. They also represent the fastest growing segment in the field of computer-aided design and computer-aided manufacturing (CAD/CAM).

ASSESSMENT OF THE FIELD

Computers allow us to routinely handle design situations whose complexity is far beyond the reach of an unaided human mind. Furthermore, the level of complexity that can be handled increases with each successive generation of computers.

For example, the computer contributed significantly to the management and design of a modern airplane such as the Boeing 767. The upper or lower wing panel on a Boeing 767 aircraft has stringers attached to it with 18,000 rivets each, the location of which is extremely critical. The wing panel is defined in the CAD/CAM system in a parametric form. More than 1,000 pieces of input information are needed to specify the exact panel location on the wing. Wing panel assembly had been a problem because it was impossible to accurately visualize the results until the stringer was essentially attached, a problem compounded by changes due to engineering error. This entire assembly was completed on the Boeing 767 using CAD/CAM. The computerized assembly allowed visualization of the entire panel prior to engineering release. It also greatly reduced error and improved fit-up so well that on the first panel assembly the actual assembly labor hours required were 72 percent less than originally estimated. In another example, the use of CAD/CAM reduced the time needed to fit the wing joint attachment from two weeks on the Boeing 747 to 18 hours on the Boeing 757.

COMPUTERS IN DESIGN

The computer extends the designer's capability in several ways. First, because the computer organizes and handles time-consuming and repetitive operations, it allows the designer to concentrate on more complex design situations. Second, it allows the designer to analyze complex problems faster and more completely. Finally, the computer allows the designer to share more information sooner with people in the company who need that design information (purchasers, tool and die designers, manufacturing engineers, process planners, etc.). These capabilities result in improved product quality, decreased design-cycle time, and designs virtually impossible to achieve manually.

The following examples illustrate these points. Using CAD, designers at General Motors were able to reduce the total weight of an average automobile by 25 percent and at the same time improve the overall structural integrity of the car. CAD reduced the time required to solve complex stress equations, allowing designers to increase the number of iterations and evaluate more completely the structural changes when metal was removed.

Using computers, GM engineers designing combustion chambers in automotive engines were able to interface a solid geometric model of the combustion chamber to a thermodynamic combustion analysis program, fire the spark, and evaluate how the flame propagates through the chamber geometry. For the first time, a designer was able to evaluate the effect of alternative geometric configurations on combustion without extensive mathematical support to set up the proper boundary conditions. Solid modeling allowed the designers to study analytical problems in a more direct and efficient way than was possible previously.

CAD systems are being used extensively in industry. Although most applications today are in two-dimensional drafting, sophisticated, but limited, three-dimensional surface and solid modeling systems are now on the market. The geometric modeling portions of these systems interface to analysis capabilities by allowing interactive problems to be undertaken. These problems are set up for finite element programs like stress analysis, for kinematics programs like mechanical linkage design, and for NC tool path generation. There is an economic trade-off between two-dimensional and three-dimensional systems, however. Three-dimensional systems require about four times more computer power than the two-dimensional systems. Companies are reluctant to acquire a three-dimensional system because they do not have the trained personnel to use it effectively. Although most CAD vendors are supplying color graphics terminals, their application appears to be for color line drawings rather than the generation of realistic color-shaded images.

Current trends in CAD are toward smaller, more powerful minicomputers. These systems have their own powerful built-in computers for stand-alone operations and can be connected to large mainframe computers for data base access or large analytical needs. Called intelligent workstations, such systems will begin to appear on engineers' desks in the near future.

Research opportunities in CAD are aimed at giving the designer more information about how his or her design will be manufactured. There are many examples in industry where different approaches to the design of the same part lead to different manufacturing procedures with widely varying costs. Future CAD systems will evaluate the merits of a design in progress and offer advice to the designer on the "best way" to proceed. Designers will also be able to access computer models of manufacturing processes for the purpose of evaluating the trade-offs of such contending processes on the manufacturability of the product.

An engineering drawing communicates the designer's ideas and requirements for a part to a machinist. The purpose of a dimension on a typical engineering drawing is to fix some distance or angle used to define an edge or surface of the part. The purpose of the corresponding "tolerance" is to give some acceptable deviation limits on that dimension. Tolerances are necessary because it is impossible to build parts with zero error or precise dimensions. Tolerances influence the machine and tool to be used, the required skill level of the operator, the expected part yield, and many other factors relating to the final cost and performance of the part.

An important step in the design of an assembly is conducting a tolerance "stack-up," or analysis, where potential clearance and interference problems that result from the buildup of tolerances are checked. Assemblies designed using CAD techniques have the advantage of an automated approach to conducting this tolerance analysis.

COMPUTERS IN MANUFACTURING

The major topics in this field are group technology, manufacturing planning and control systems, automated materials handling, computer-aided manufacturing, and robotics.

Group Technology

Group technology is the classification and coding of parts (material, size, shape, volume) and processes (lot size, machine tools, setup time) into families with similar characteristics. It enables the designer to evaluate designs to determine which parts already exist, which parts can be modified from existing parts, and which parts are truly new. Group technology facilitates the establishment of manufacturing cells (groups of machines coordinated to create a given part) for a family of parts, and provides a basis for computer-aided process planning. Direct benefits include a smaller number of parts in inventory (and in the data base), reduced time to design a new product, and increased use of capacity.

Group technology plays an important role in the planning function. Consequently, research is needed to develop models for manufacturing cells or hierarchies of cells that can be used to generate overall optimal scheduling strategies.

Manufacturing Planning and Control Systems

Manufacturing planning and control generally include the following functions: material resource planning (MRP), distribution resource planning, overall simulation and optimization, and shop floor data collection. These functions represent the management and business information needs of the entire enterprise.

The system developed over the last eight years by Ingersoll Milling Machine Company, a large discrete-parts job shop, illustrates the state of the art.* That system integrates a management and business information system with a CAD/CAM system. The system's master scheduler links the bill-of-materials and geometry files to coordinate dispatch, inventory, MRP, planning, and scheduling. High-level management decisions are automatically entered in the system, and each supply source (machine shop, purchasing, stockroom) receives only the appropriate information.

This system has expanded dramatically the existing system's design and implementation capability and provided a major new role in strategic planning. Within four years the productivity of the system's design activity doubled. Today, 82 percent of the system's design time is spent in creative development and only 18 percent on routine maintenance (compared to 42 percent spent in creative development four years old).

The general trend in the area of manufacturing planning and control is toward the integration of more business planning and reporting functions. Although simulation techniques are used extensively to test scheduling scenarios, they cannot yet generate them. An important research problem is the refinement of large-scale optimization techniques to identify planning and scheduling strategies for large, hierarchical, flexible manufacturing systems. Other important topics include inventory management techniques to accomplish "just-in-time" objectives and the employment of knowledge-based expert technologies to ensure quality throughout the design, test, and manufacturing cycle.

Automated Materials Handling

The major automated materials handling systems consist of automated storage and retrieval systems, miniloaders, automated guided vehicle systems, and conventional materials/parts storage and transport systems. Benefits include higher inventory record accuracy, reduced storage space requirements, increased labor productivity, and the automatic transmittal of material movement information from the production control system to materials handling equipment.

*The company received the 1982 LEAD award (Leadership and Excellence in the Application and Development of Computer-Integrated Manufacturing) from the Computer Automated Systems Association of the Society of Manufacturing Engineers.

Computer-Aided Manufacturing

Computer-aided manufacturing generally refers to numerically controlled (NC) machining and other processes (including flexible manufacturing systems), computer-aided inspection and testing, automated assembly, and process control.

In the machining area, NC machines have been used for many years with proven benefits in the form of reduced time for machining and storage. The current trend in the industry is a logical extension of NC to 2 to 15 machine tools linked with automated materials handling and production scheduling—a machining cell. Such systems are commonly used to produce one or more families of parts. Benefits provided by these systems include increased production flexibility, reduced operating costs, and higher machinery utilization.

Computer-aided inspection is an area of intense research. The benefits include automatic collection and analysis of quality control data, creation of data bases to isolate production process problems, and correction of machinery problems before many faulty parts are produced.

A system developed at General Electric Corporate Research and Development for inspecting jet engine turbine blades represents the state of the art. The basic task for that system is to inspect turbine blade surface and subsurface geometry to exacting tolerances at high rates. The inspection system consists of three major subsystems: a subsystem to present and sense the part precisely, a signal-processing and image-processing subsystem to inspect the part, and an analysis and decision-making subsystem to accept or reject the part automatically.

In the electronics area, IBM has developed a robotics-based automated system to test the thousands of connections on the electronic boards of mainframe computers. This system is based on design data that describe the location of the connecting paths.

Automatic assembly systems today are generally built to perform one task. The trend in this area is toward flexible, electronically controlled systems, including assembly robots. It is possible to generate robot programs automatically for well-defined and regular geometric layouts such as circuit boards and cable harnesses.

Model-based CAM has the capability to generate computer programs automatically from geometry and other information, in order to control manufacturing processes and to simulate such processes for verification and evaluation. Automatic procedures would extract pertinent information from underlying models of various processes such as extruding, cutting, forming, casting, welding, painting, hardening, testing, assembly, and material flow. Mathematical bounds on the reliability and optimality of these models are also needed.

In primary processes, such as molding or forming, it is possible to derive mold geometry from part geometry. Modeling mold performance is still largely a research issue.

Logistics models involve material flow and process planning. From a network flow model it is possible to predict assembly-line behavior automatically. However, there is limited ability to derive optimal schedules. It is possible to derive process plans for the sequence of manufacturing steps required for axially symmetric parts.

Research is needed in modeling almost all processes. Models are also needed for sensor-mediated assembly and visual inspection, i.e., the use of sensors to control the process. Such models require better methods of modeling three-dimensional geometry and constraint relationships, including kinematics and tolerances. Methods are needed to permit fabrication planning from three-dimensional geometry.

Robotics

Robotics is an area of intense activity. The major industrial applications include spot welding, materials handling, finishing and coating, arc welding, and assembly. The two most promising application areas for future growth are materials handling, including loading and unloading operations, and assembly.

Current robots can offer very precise and repeatable position control. However, the availability of force sensing, the ability to control the force exerted by the robot's grasping mechanism, is very limited. Virtually no continuous force control is available.

Programming systems are somewhat incompatible with most standard operating systems and programming languages. They also lack features such as coexistence of interpreted and compiled code, configurability, and facilities for debugging, multitasking, multiprocessing, and networking. Smart (adaptive) programming systems are virtually nonexistent.

Current robot vision capabilities are two-dimensional binary and are generally restricted to well-separated parts. Development is under way on applications and systems for structured-light, three-dimensional systems (i.e., systems using light projected in coded form), which will greatly expand the capability.

Given current research, one can expect to see in the next several years general purpose, real-time distributed operating systems and programming languages. Robots will exhibit limited force control, assembly capabilities, and coordinated use of multiple arms. There will also be commercial, structured-light, three-dimensional systems, and limited model-based vision systems.

Methods are needed for linking robots and vision systems to CAM models to allow real-time use of such models. Research is needed on establishing a set of fundamental low-level commands to characterize sensor-mediated actions, such as force-controlled motion.

INTEGRATION OF THE TOTAL SYSTEM

The total integration of all computer-based systems that support design, manufacturing, and business functions is essential to realize the potential of the computer for increasing productivity. Although significant inroads have been made on a limited scale in specific industries, system integration on a large scale is an extremely complex and poorly understood problem.

The requirements of an integrated system include (1) analytical, operating, and control systems that communicate directly with each other; (2) systems able to generate (by creating, deriving, transforming, and interpreting data) the appropriate information necessary for the support and operation of the design and manufacturing functions; (3) the use of a common data base concept that can be physically distributed over many computers; (4) a responsive communications network; and (5) a master system to manage the whole.

Research is needed to understand the nature of the integration problem. We expect that simulation techniques will be useful in modeling integration concepts to help evaluate the anticipated effects, the reliability, the usefulness (cost/benefit), and the potential likelihood of achieving large-scale integration. Research on this problem is risky because of the comprehensive nature of the changes involved and the enormous investment required. Yet it carries the greatest potential for payoff.

Global Information Structures

We can think of existing research in design and manufacturing as creating the building blocks of a cathedral. We are putting the blocks in place to build the cathedral without actually knowing if the overall structure is going to stand up.

Currently, data carried in traditional industrial data bases is functionally divided to support engineering and manufacturing separately. The engineering data base supports geometry and engineering business data, such as scheduling and purchasing. The manufacturing data base supports the processing aspects of geometry, such as numerical control, and manufacturing business data such as planning and inventory control. In most cases these data are located in several physical data bases.

The data interfaces must be structured to support all the above industrial processes. The structure must be flexible enough to accommodate the information needs for different applications such as machining, scheduling, inventory control, estimating costs, analysis, or design. It is also desirable that the information be derived from the data without human intervention. Yet it must be general enough to permit the integration of current systems and still be extendable to future systems. Such an information structure will in effect become the master source for the definition of both product and process.

The nature of global information structures is not well understood today. Data flow concepts need to be developed and tested against a broad spectrum of applications.

Distributed Data Management and Networking

Applications requiring large data bases and involving different computers in a network create a severe communication problem because most data base management systems run on only one computer. Consequently, data communication problems arise because the data base management system in one computer cannot be used to query data in another computer.

Long-distance data communication between separate geographic sites is another typical problem in heterogeneous computer networks. Current services for the transmission of critical geometric and other data are slow, noisy, and costly. Efforts to establish standards are proceeding but are still unresolved.

Integration of the design and manufacturing of a product can occur only if the heterogeneous software and hardware systems support a common distributed data base. This distributed data base must have adequate recovery and backup procedures when simultaneous updates on different computers occur or when one or more computers in the network fail. It must have minimal redundancy of data while maintaining the most efficient access route and data traffic. Further, it must maintain audit trails of a large variety of simultaneous updates caused by design or manufacturing changes. Research is needed to study the data characteristics necessary to integrate the design and manufacturing systems within a distributed data base environment composed of heterogeneous computers.

RESEARCH OPPORTUNITIES

The trends in the application of computers in design and manufacturing identified in the previous section suggest a number of research opportunities. Underlying these areas are a set of generic research topics that can be grouped into the five following areas: (1) modeling, (2) human-computer interface, (3) expert systems, (4) information management, and (5) devices. Research in these areas will provide the critical building blocks needed to speed the application of computers in the factory.

MODELING

Geometric models embody the description of a product design. Consequently, they represent a basic link between CAD and CAM. Most existing CAD systems describe product geometry by means of potentially ambiguous "wireframe" representations, a collection of points and lines. With wireframe representations there is no way of determining automatically if a point was "inside" or "outside" the object. It is not possible to calculate mass properties (i.e., volume, inertia) in any automatic way. Therefore, automated analysis (and model-driven CAM) will require a complete and unambiguous representation of the geometry in three dimensions.

Geometric Modeling

The ultimate goal of geometry research is to develop a complete and unambiguous three-dimensional geometric representation for part and process modeling. This representation should include free-form (sculpted) surfaces, allow creation of geometry in a natural manner (be user friendly), and permit automatic analysis and process planning.

Current Status The theoretical foundations for computer-based solid geometry representations were laid through university research conducted primarily in the U.S. and U.K. during the 1970s. U.S. industry is just beginning to use solid modeling in production. The solid modelers that are commercially available all have shortcomings.

Industrial research that is focused on methods for improving existing techniques should lead in a few years to solid modelers acceptable to industry with increased capabilities in surface sculpting and blending. Performance should also increase, primarily because of new hardware and improved software.

. . .

Key Research Needs An important research problem is the integration of free-form surface generation capabilities into solid modelers. Several approaches currently being investigated appear promising, including the rational B/SPLINE representation and the generation of blending surfaces between a given set of primitive solids. The deformation of solids is another important problem area where very few inroads have been made. The ability to handle deformations, for example, has very important practical implications in modeling the extrusion process. One can identify many such process operations that need to be investigated.

Another research area involves tolerancing information for solids. Current mathematical and computational theories of mechanical tolerancing have significant gaps. Finally, there are many geometry problems such as surface-surface intersection, stable position of free-form solids, numerical accuracy of geometric models, and new geometric primitives that need to be investigated.

Analytic Modeling

There are many analyses needed in the course of a design that involve the creation of an approximate (discretized) geometric model and then the application of loads and material properties. Thermal and stress analyses using finite element techniques are two examples.

Two steps are required to perform these analyses automatically. First, we need an automatic discretization procedure that will provide some type of approximate form of the initial geometric object. Second, we need a way of ensuring beforehand that the numerical results of the analysis will be within a prescribed level of accuracy (adaptive analysis).

Current Status Research is needed in analysis methods that can be applied directly to a geometry configuration with associated boundary conditions in some automatic way. For example, research on automatically generated meshes for finite element analysis is a very active area. Industry-usable results should be forthcoming in the next several years.

Key Research Needs Adaptive analysis is the primary area of needed research. We need the capability of applying an analysis procedure such as stress analysis directly to a geometric model with given boundary conditions and automatically obtaining an answer to a specified level. This problem has a high research priority because it represents an important step in the automation process. Expert systems are needed to aid the designer in choosing an appropriate approximation model for each condition in the problem (e.g., shell model or beam model for different parts of an airplane).

HUMAN-COMPUTER INTERFACE

The human-computer interface in CAD is the link between a design engineer and the CAD system. A user-friendly interface will guide the designer through complex design tasks in a direct and effective manner. This is very important, especially in the early design decisions that have a major impact on the total design-production process. The development of such an interface requires an understanding of the creative design process. Ideally, the system will act as a natural extension of the designer. Friendly interfaces are especially important in CAM, where users may be less familiar with computing techniques.

Current Status Graphics systems are now very sophisticated, with the ability to create color-shaded, three-dimensional images from a design data base. However, most existing human interface capabilities are still relatively crude with very limited help capabilities to get out of trouble. In short, today's occasional user faces an imposing learning experience each time he or she sits down at the workstation. Graphics and engineering workstation vendors are beginning to pay attention to the relevant human factors and human engineering issues, but the effort is still very limited.

Key Research Needs Research is needed to understand better how an engineer conceptualizes, since it is in the early stages of design that current CAD/CAM is least effective. Modeling cognitive processes associated with design need to be explored. Research is needed in developing "smart" CAD systems that use expert (i.e., knowledge-based) concepts to guide the engineer interactively through a design situation by continually evaluating his or her performance and issuing guidance in terms of the individual's perceived experience base.

EXPERT SYSTEMS

Expert systems consist of a body of knowledge and a mechanism for interpreting this knowledge. The body of knowledge is divided into facts about the problem and heuristics or rules that control the use of knowledge to solve problems in a particular domain.

In the last few years, expert systems have become the most visible and fastest growing branch of artificial intelligence. Their objective is to capture the knowledge of an expert in a particular area, represent it in a modular, expandable structure, and transfer it to other users. To accomplish this goal, it is necessary to address issues of knowledge acquisition, knowledge representation, inference mechanisms, control strategies, user interface, and finally how to deal with uncertainty.

Current Status Expert systems development, confined during the past decade to academic laboratories, is now becoming commercially viable, partly because of the development of well-understood methods for knowledge-based programming and partly because of advances in microelectronics. Current examples of expert systems include MOLGEN, which interactively aids molecular geneticists in planning DNA-manipulation experiments; VM (Ventilator Management), which gives real-time advice for the management of patients undergoing mechanical ventilation in an intensive care unit; PROSPECTOR, which advises when and where to drill for ore; and DELTA, a production rule-based system for diesel electric locomotive repair. Expert systems have not yet been applied in any real sense to solve manufacturing problems.

Key Research Needs Applied research is needed to gain some experience in building knowledge bases for selected manufacturing processes, especially in planning functions. Research is needed to sort out the issues of "design for manufacturability" before undertaking the task of building one. The concept of geometric reasoning is very complex; we need to understand how to characterize manufacturing "reasoning" such as "find a symmetrical object in the data base." Finally, we need to examine closely the issues or experiences involved in developing knowledge bases for making approximations in complex analysis problems.

INFORMATION MANAGEMENT

This area concerns the management of design and manufacturing data so that users and systems can use the data easily and efficiently.

A data base is an organized collection of discrete data on a given subject. A data base is made accessible and useful by means of a data base management system (DBMS). This management framework permits users to both query and update the data base in order to derive comprehensive lists of data having certain characteristics in common.

. . .

Current Status Data base management systems are commercially available for a wide range of applications. Current research is producing considerable advances in DBMS capabilities, including improved performance, modeling capabilities; and user interfaces. Research on distributed data bases is a critical area. However, there are features of the manufacturing environment that render current data bases and DBMSs inadequate. We do not know how to make data base systems that efficiently handle manufacturing requirements in a distributed environment.

Key Research Needs Research is needed to organize geometric data in a way that allows "geometric reasoning," i.e., to optimize the data extraction operations in response to continually developing and unanticipated user queries. For example, consider requesting a list of parts where the parts possess a certain symmetry, regardless of the sequence of geometric steps used to construct the parts originally. We may not know beforehand that symmetry is important and therefore would not want to assign a symmetry attribute. We want the system to "reason" its way to the correct information.

Other research areas include the following:

- A basic study into the nature of the integration problem. This could lay the groundwork for integrated experimentation on a larger scale.
- The development of data characterization requirements for the integration of systems in a distributed data base environment.
- The investigation of the information structure required to support the so-called neutral data base or master product definition.
- The development of algorithms to support proper recovery and backup procedures of distributed data bases in heterogeneous computer environments.
- The use of simulation techniques to help model integration to understand better the anticipated effects, reliability, usefulness (cost benefits), and the overall potential to achieve actual integration.

DEVICES

This section deals with issues at the device or workstation level as opposed to systems level. The basic technology of robot sensors and actuators must be advanced significantly in order to achieve the needed order-of-magnitude improvement in precision, speed, and dexterity. Component modularity involving new micromechanical, electronic substrate-based sensors and controllers is required for the flexibility needed to create unique manufacturing systems. The ability to integrate and coordinate comprehensive strategies for sensing and control is also very important.

These objectives imply the need for special-purpose computer architectures and VLSI (very large scale integrated) chip technology. Such chips will also be essential for handling the growing computational workload in automated CAD and CAM modeling and analysis operations. Chips for handling specialized geometric constructions, inference interpretations, and analysis processor calculations will dictate requirements for special-purpose microelectronic chips in the near future.

Current Status Currently, we have a heterogeneous collection of sensors and actuators and no uniform way of integrating these components into a coordinated manufacturing system. For example, vision systems are important not only for robots but also for all kinds of inspection and control functions.

Array processors for special-purpose analysis calculations and display processor chips for graphics are becoming more widely used. They are only the beginning of a trend toward fast hardware.

Key Research Needs Research is needed to develop smart sensors, including visual and tactile sensing. The investigation of modularity issues in sensors and actuators is also important. Finally, the development of requirements for special-purpose processors to implement advanced algorithms in analysis, geometry, expert systems, network communications, and control strategies are needed. The research issues here cut across other fields in a significant way. Collaboration will be required. This research is important because the ultimate precision and speed of our manufacturing will be dependent on the inherent capabilities of these devices.

THE ROLE OF UNIVERSITIES, INDUSTRY, AND GOVERNMENT

Universities, industry, and government traditionally have been the key institutions in manufacturing technology research. Each has its own method of addressing research problems, and each has unique limitations and strengths.

SUPPORT MECHANISMS FOR RESEARCH

In universities, research tends to be small scale in nature, where specific, well-formulated problems are studied in a formal manner. Universities have been reluctant to grapple with the larger problems of integration, partly because of inadequate interdisciplinary knowledge (few faculty trained to address such issues) and because of an uncertainty about the scientific issues. Hardware and facilities are costly—usually beyond the means of a university budget. Even when hardware is donated by industry, the operational costs are substantial. In addition, many graduate students in computer-related fields are attracted by industry, sometimes before completion of their degree program. This effectively limits the formation of a cadre of qualified researchers. Government and other funding for university research in this field is insufficient to develop the critical mass of researchers needed to sustain significant progress.

Industrial companies have generally solved design and manufacturing problems piecemeal, addressing short-term objectives. Broad gaps in basic knowledge are increasingly apparent as computers are applied to factory problems that pose larger and more difficult integration hurdles. Compounding these organizational limitations are some situational problems.

1. It is difficult for industry to articulate its systems' integration problems. The issues vary greatly by company size and markets.
2. Manufacturing equipment is generally tied up for production purposes and unavailable for scientific experimentation.
3. The industrial application of research findings requires capital and personnel investments over a period of years, with accompanying high costs. Also required are in-house advocates to sell the concept, lead the implementation, and provide a critical mass of management approval.

The federal government has been a major customer for existing CAD/CAM products and for highly specialized, one-of-a-kind systems. This latter role has catapulted key government agencies (notably the U.S. Department of Defense and the National Aeronautics and Space Administration) into areas of systems design and management, so that they have become catalysts for advances in knowledge through acquisition, internal research, and sponsored research.

Clearly, our view of these problems is at a point of transition.

During the past decade agencies of the federal government have embarked on several programs that have begun to focus attention on both component development needs (the National Bureau of Standards and the U.S. Department of Defense) and system integration problems (the U.S. Air Force and the National Aeronautics and Space Administration). An important factor in these programs is the use of consortia of industrial companies and universities, although university involvement has been limited.

There are a few recent examples of university/industry collaboration at schools such as Rensselaer Polytechnic Institute, Carnegie-Mellon University, and Stanford University. Other universities are attempting to develop similar collaborative efforts, many of them in local areas. It appears that an effective way to conduct research and transfer technology in certain areas would be via these consortia.

Over the years, individual companies have supported applied and generic research at universities through grants of money and equipment. This activity has now been intensified, the most recent development being IBM's announcement in July 1983 of $10 million to five universities for manufacturing-related curriculum development. This was preceded by a $40 million equipment grant to 20 universities.

Individual companies within industries are beginning to explore the feasibility of creating joint R&D partnerships patterned after Microelectronics and Computer Technology Corporation, which was created within the electronics industry to undertake the study of generic research issues.

The central issue in the field of computers in design and manufacturing is a pervasive lack of scientific knowledge. This lack of knowledge is less pervasive than it was 10 years ago, in part because university research has been able to make some significant accomplishments with very limited resources. However, major gaps still exist, as this report illustrates. The first priority of this field is to build a genuine, classical research community. The large integration efforts identified in this report will require a much stronger base of fundamental knowledge than is now available.

EDUCATION AND TECHNOLOGY TRANSFER

As important as the development of technology is the transfer of that technology to an educated work force (including engineers, managers, and operators). There are four basic issues, each requiring a different mechanism. The first issue is the transfer of prototype technology from

R&D laboratories to industrial use. University-industry consortia can facilitate this transfer, as will the traditional mechanisms of conferences, publications, and product sales.

The second problem is to educate a new breed of engineers who thoroughly understand all aspects of computer-integrated manufacturing (CIM) engineering in its broadest sense. This involves developing new courses and programs and implies a knowledgeable faculty.

The third issue is the integration of modern computing tools into the traditional program of engineering education. These students will become the users (not developers) of CIM.

The last issue is the training of operators and technicians in an effort to upgrade the existing work force. It is necessary to find cost-effective ways of training people at a significantly faster rate and to develop a means of transferring experience without incurring the high cost of one-on-one instruction. Better productivity in training for skill development is a critical need.

RESEARCH BRIEFING PANEL ON SELECTED OPPORTUNITIES IN CHEMISTRY

George C. Pimentel ***(Chairman),*** *University of California, Berkeley*

Allen J. Bard, *University of Texas*

John H. Birely, *Los Alamos National Laboratory*

Michel Boudart, *Stanford University*

Ronald Breslow, *Columbia University*

Harry B. Gray, *California Institute of Technology*

Ralph Hirschmann, *Merck Sharp & Dohme Research Laboratories*

Alan Schriesheim, *Exxon Research & Engineering Company*

Howard E. Simmons, *E. I. du Pont de Nemours & Company*

Barry M. Trost, *University of Wisconsin*

George M. Whitesides, *Harvard University*

In cooperation with the NRC Committee to Survey the Chemical Sciences and the Board on Chemical Science and Technology

Staff

William Spindel, ***Executive Secretary,*** *Board on Chemical Science and Technology (BCST)*

Mary E. Bundy, ***Administrative Associate,*** *BCST*

Peggy J. Posey, ***Staff Associate,*** *BCST*

Robert M. Simon, ***Rapporteur***

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON SELECTED OPPORTUNITIES IN CHEMISTRY

EXECUTIVE SUMMARY

This is a time of special opportunity for intellectual advance in chemistry. It derives from our developing ability to probe and understand the elemental steps of chemical change and, at the same time, to deal with extreme molecular complexity. Powerful instrumental techniques are the crucial dimension. They account for the recent acceleration of progress that gives chemistry unusual promise for high return from additional resources.

Chemistry is a central science that provides fundamental understanding needed to deal with most societal needs: to feed the world population, to tap new sources of energy, to clothe and house humankind, to provide renewable substitutes for dwindling or scarce materials, to improve health and conquer disease, to add to our national security, and to monitor and protect our environment. Further, there can be no doubt that chemistry, with its current $12 billion positive balance of trade, is a crucial element in this nation's economic well-being. Our future international competitiveness will depend upon maintaining our present leadership position in the chemical sciences.

Directions of particular promise are becoming apparent in the course of the NAS/NRC study by the Committee to Survey the Chemical Sciences. This Committee is formulating a multi-year program to exploit the richest potentialities in a priority sequence that will maximize return on the needed additional federal investment. We identify here three areas, those that will yield the greatest rewards and furnish the appropriate beginning to this program.

A. Understanding Chemical Reactivity
B. Chemical Catalysis
C. Chemistry of Life Processes

A. Understanding Chemical Reactivity

We propose an initiative to apply the full power of modern instrumental techniques and of chemical theory to the clarification of factors that control the rates of reaction and to the development of new pathways for chemical change.

- *Molecular Dynamics:* to elucidate the entire course of chemical reactions, including the unstable atomic arrangements intervening between reactants and products.
- *Reaction Pathways:* to accelerate our rate of discovery of new reaction paths and new classes of compounds.

A principle objective of this initiative will be to provide the fundamental basis for U.S. leadership in development of new processes, new substances, and new materials.

B. Chemical Catalysis

We propose an initiative to apply the techniques of chemistry to obtain a molecular-level understanding of catalysts. Ultimately, new catalyst systems will result that will lay the foundation for the development of new chemical technologies. The program will emphasize research in four areas:

- *Heterogeneous Catalysis:* to apply the powerful new instrumental techniques of surface science to the study of chemistry on the surface of a solid.
- *Homogeneous Catalysis:* to take advantage of recent developments in synthetic chemistry that show promise as new soluble catalysts.

- *Photocatalysis and Electrocatalysis:* to investigate solution chemistry catalytically assisted by electrode processes, with and without absorption of light.
- *Artificial-Enzyme Catalysis:* to couple the chemists' ability to control molecular topography with the biochemists' understandings of natural catalysts so as to generate a new class of artificial enzyme-like catalysts tailored to specific needs.

A principal objective of this initiative will be to provide the fundamental knowledge and creative manpower required for the U.S. to maintain competitive advantage in catalysis-aided technologies.

C. Chemistry of Life Processes

We propose an initiative to develop and apply the techniques of chemistry to the solution of molecular-level problems in life processes. Research at this border between chemistry and biology requires individuals broadly competent in both areas, and a special effort must be made to develop such individuals. The program will emphasize the molecular aspects of six areas:

- Enzymology (understanding the molecular interactions responsible for enzymatic activity; production of natural enzymes for use as catalysts in chemical synthesis).
- Immunochemistry (the chemical basis of the immune and allergenic response; use, function, and modification of monoclone antibodies; synthesis of antigens and adjuvants).
- Chemical Endocrinology (synthesis of hormones and hormone analogs, especially those related to polypeptides).
- Neurochemistry (molecular basis of nerve transmission; neurotransmitters; agonist and antagonist chemistry; membrane polarization).
- Membrane Chemistry and Vectorial Chemistry (bioenergetics; active and passive transport).
- Biological Model Studies (host/guest chemistry; semisynthetic enzymes; properties of aqueous solutions; active site modeling).

A principal objective of this initiative will be to accelerate the conversion of qualitative biological information into techniques and substances useful in biotechnologies, in human and animal medicine, and in agriculture.

In the next two decades there will be dramatic changes in our basic understandings of chemical change and our ability to marshal those understandings to deliberate purpose. The program presented here is intended to define a leadership role for the U.S., as these advances are won. The rewards accompanying such leadership are commensurate with the prominent role of chemistry in addressing societal needs, in ameliorating the problems of our technological age, and in sustaining our economic well-being. The costs of falling behind are simply intolerable.

The need for enhanced federal investment in chemical research is rooted in a pattern of funding historically appropriate to a test-tube and bunsen burner era, an era long since eclipsed. The sophistication of a modern chemistry laboratory requires a much more vigorous financial commitment, both in capital investment and in supporting services. Even so, the cost is miniscule compared to the stakes involved; we must nurture a $175 billion industry that maintains a $12 billion positive balance of trade. We must provide it with a full reservoir of fundamental knowledge and we must attract a substantial share of our brightest young scientists to this rewarding task. This program can accomplish these goals.

I. INTRODUCTION

This is a time of special opportunity for intellectual advance in chemistry. It derives from our developing ability to probe the elemental steps of chemical change and, at the same time, to deal with extreme molecular complexity. Powerful new instrumental techniques are the crucial dimension. They account for the recent acceleration of progress that gives chemistry unusual promise for high return from additional resources.

Chemistry is a central science that provides fundamental understanding needed to deal with most societal needs. It is a critical component in man's attempt to feed the world population, to tap new sources of energy, to clothe and house humankind, to provide renewable substitutes for dwindling or scarce materials, to improve health and conquer disease, to add to our national security, and to monitor and protect our environment. Basic research in chemistry will surely help future generations cope with their evolving needs and unanticipated problems.

Chemistry is also a crucial element in this nation's economic well-being. The U.S. Chemical and Allied Products industry employs over a million people, makes manufacturing shipments totaling about

$175 billion, and currently displays a $12 billion positive international balance of trade, second highest of all commodity groups. Our competitiveness in a range of international markets depends upon maintaining our present position of leadership in the chemical sciences. There is no area of basic science that offers a more secure investment in the nation's future.

Thus, rapid advances are possible now in chemistry, advances that will continue to enrich our cultural heritage, that will ultimately respond to human needs, and that will sustain our economic competitiveness. Directions of particular promise have become apparent in the course of the NAS/NRC study by the Committee to Survey the Chemical Sciences. This Committee is formulating a multi-year program to exploit the potentialities in a priority sequence that will maximize return on the needed additional federal investment. We identify here three areas that furnish the appropriate beginning to this program.

A. Understanding Chemical Reactivity
B. Chemical Catalysis
C. Chemistry of Life Processes

II. PRIORITY OPPORTUNITIES

Three priority thrusts are described. The first is connected with opportunities to understand, in the most fundamental sense, chemical reactivity and to control it. The second is to advance our understanding of catalysis in all of its facets. The third will extend to the molecular level our understanding of life processes. These three thrusts are complementary and reinforcing in the sequence given. They have different rationals and distinct budget needs, so they are presented separately.

A. Understanding Chemical Reactivity

1. Program.

We propose an initiative to apply the full power of modern instrumental techniques (lasers, molecular beams, computers, etc.) and of chemical theory to the clarification of factors that control the course of a chemical reaction and to the development of new pathways for chemical change.

- *Molecular Dynamics:* to elucidate the entire course of chemical reactions, including the unstable atomic arrangements intervening between reactants and products.
- *Reaction Pathways:* to accelerate our rate of discovery of new reaction paths and new classes of compounds.

2. *Objectives.*

- *To advance our fundamental understanding of chemical reactivity.*
- *To attract exceptional young scientists to the frontiers of this understanding.*
- *To apply this understanding to the development of new types of reaction and new classes of compounds.*
- *To provide the fundamental basis for U.S. leadership in development of new processes, new substances, and new materials.*

3. *Rationale.*

Public support for basic research is justified, in part, on cultural grounds and, in part, on practical grounds. Our culture includes the premise that learning about ourselves and our environment is an ample basis for encouragement of scientific investigations. In addition, we recognize that our increasing ability to meet the common desire for survival, comfort, health, and freedom from toil is rooted in our increasing comprehension of what goes on within and around us.

Within the cultural basis for science, three prime questions stand out: the structure of the Universe, the makeup of matter, and the mechanism of life processes. The third is probably the strongest human preoccupation: the nature and preservation of life. Since all life processes—birth, growth, reproduction, aging mutation, death—are manifestations of chemical change, understanding chemical reactivity can be seen as the foundation for our ultimate understanding of life. Thus chemistry, along with biology, can properly be placed with astronomy and high energy physics as scientific investigations justifiable because they contribute to human knowledge in areas of universal philosophical significance.

However, that does not complete the justification for increasing our understanding of chemical reactivity. It is just as convincingly based upon practical grounds. The development of new processes and new materials depends upon our understanding and control of chemical change. Fundamental advances here will present avenues for innovation that can be exploited in the study of combustion, corrosion and electrochemistry, polymer formation, tailored organic molecules, and new solid state materials. It responds to our need to develop new energy sources, to provide alternate materials for non-renewable or scarce materials, and to maintain economic competitiveness, all while attentively protecting our environment. It permits the U.S. to produce billions of pounds of organic chemicals, at low cost, in high yield, and with minimum waste product. Combined with our 1981 production of 9.8 billion pounds of synthetic fibers (such as polyesters), 28 billion pounds of plastics (such as polyethylene) and 4.4 billion pounds of synthetic rubber, this capability accounts for the total business volume mentioned earlier, an amount exceeding $175 billion with its current international trade surplus of $12 billion.

This is surely a time of special opportunity to deepen our fundamental knowledge of why and how chemical changes take place. The advance of the frontiers of molecular dynamics has undergone a revolutionary acceleration during the last decade. At the same time, synthetic chemists are exponentially increasing the number of known molecular structures, and constantly adding to our arsenal of reaction types and classes of compounds. Much of this can be attributed to the development and application, during the last two decades, of an array of powerful instrumental and analytical techniques. A few examples show that we have new capabilities in hand with which to probe far beyond current frontiers.

MOLECULAR DYNAMICS

Lasers by themselves have spectacularly expanded our experimental horizons over the last decade. Their short pulse durations permit easy probing of chemical reactions that occur in less than a millionth of a second down to times a thousand-fold shorter (from a microsecond to a nanosecond). With more complex (and more expensive) equipment, chemists are now entering a time domain still another thousand-fold shorter (to a picosecond). Lasers also provide tunable, extremely narrow frequency light sources, implying vastly greater diagnostic sensitivity and selectivity. Finally, lasers give us extremely high light intensities, either in the form of short pulses of ultrahigh power or as continuous light sources with unprecedented photon flux.

Computers also make important contributions to this revolution, both in experimental and theoretical chemistry. They are now common components of experimental assemblies and a number of the chemist's more sophisticated instruments have built-in (dedicated) computer capability as an essential element. (For example, fourier transform infrared spectrometers, fourier transform nuclear magnetic resonance spectrometers, and X-ray spectrometers). Furthermore, today's computing capability has brought *ab initio* calculations and chemical theory to new levels of fruition.

These are only exemplars—a host of sophisticated instrumental methods ought to be mentioned: molecular beams, ion cyclotron resonance, electron spin resonance, photoelectron spectroscopy, magnetic circular dichroism, resonance Raman, Mossbauer, etc. All of these account for the rich possibilities lying before us. However, the historical basis for funding basic research in chemistry was not premised on the existence (and cost) of such sophisticated equipment. One of the strongest arguments for a large increase in federal support for chemical research is that existing levels do not provide sufficient access to these state-of-the-art techniques, and without them U.S. science will fall behind in areas crucial to our national well-being.

Turning to these new frontiers, the advance of understanding of *molecular dynamics* is one of the most rapidly moving. Here are some of the new phenomena we can now explore.

- *Fast chemical process in real time.* The primary processes of numerous chemical changes are determined by dynamical events that are intrinsically fast. Many solution phase reactions, photochemical reactions, electron solvation, molecular reorientation in liquids, and reactions of certain biological systems can now be watched "in real time" on the nanosecond-to-picosecond time scale.
- *Energy transfer and movement.* In all chemical changes, the pathways for energy movement are determining factors. Competition among these pathways (including energy dissipation) determines the product yields, the product-state distributions, and the rate at which reaction proceeds. Chemists are now able to track these pathways: radiation, internal conversion, intersystem crossing, intramolecular and intermolecular energy transfer, unimolecular decomposition.
- *Ab initio calculations of reaction surfaces.* With today's computers, the structure and stability of any molecular compound with up to 3 first-row atoms (carbon, nitrogen, oxygen, fluorine) plus various numbers of hydrogen atoms can be calculated almost to the best accuracy available through experiment. This opens to the chemist many situations not readily accessible to experimental measurement. Short-lived reaction intermediates, excited states, and even saddle points of reaction can now be understood, at least for small polyatomic mole-

cules. In a major advance, we can now calculate the forces on all of the atoms during their reorganization from reactant to product molecule geometries.

- *State-to-state chemistry.* Major developments in molecular beam technology, including mass spectrometric ("universal") detectors, vacuum technology, laser excitation, and supersonic jet sources, have enormously broadened their applicability. In the ultimate experiment, reactants can be brought together in precisely known energetic states, with every quantum number fixed, and then the product energies and relative probabilities can be established in the same detail.
- *Mode-selective chemistry.* With high-power, sharply tunable lasers, it is possible to excite one particular degree of freedom for many molecules in a bulk sample. As long as this situation persists, such molecules react as if this degree of freedom is at a very high temperature while all the rest of the molecular degrees of freedom are cold. The chemistry of such selective excitation could potentially reveal the importance and role of that particular degree of freedom in facilitating reaction. To extract this valuable information, energy redistribution and relaxation must be brought under control, and this frontier is being explored.
- *Multiphoton excitation.* One of the most surprising discoveries of the 1970's was that powerful pulsed lasers could highly excite the vibrational degrees of freedom of a molecule on a time scale short compared to molecular collision times. With many tens of quanta of vibrational excitation, the molecule displays unusual chemical pathways. This opens a new type of photochemistry with unique possibilities, the first of which to be recognized was isotope enrichment.

The rich potentialities of these techniques promise to show us how reactions proceed over their entire course. We will at last be able to understand the unstable situations intervening between reactants and products. To realize these benefits, there must be wider access to the sophisticated instrumentation (including computing capability), and the cadre of bright young scientists attracted to this area must be encouraged and increased.

REACTION PATHWAYS

A manifestation of this increasing understanding and control of chemical reactivity is the rapid advance now taking place in devising new reaction pathways in synthetic chemistry. This progress presents a high leverage opportunity since herein lies the foundation of future developments of new products and new processes.

Again powerful instrumental techniques play a central role. The rapid and definitive identification of reaction products, both in composition and structure, accounts for the speed with which synthetic chemists are able to test and develop their synthetic strategies. The nuclear magnetic resonance and mass spectra show what elements are present and the structural environment of virtually every atom. The X-ray crystal structure reveals the complete molecular structure: the interatomic distances, bond angles, and even the right- or left-handedness when mirror-image relationships are present. Spectroscopic techniques, as well, have been essential to progress in the rapidly developing area of organometallic chemistry. The visible and infrared spectra of transition metal complexes reveal electronic configurations and bonding, the foundation for clarifying mechanisms for ligand substitution and electron-transfer processes.

■ *Selective pathways in organic synthesis.* Selectivity is the key challenge to the synthetic chemist—to make a precise structural change to a single desired product. The different intrinsic reactivity in each bond type must be recognized (chemoselectivity), reactants must be brought together in proper orientation (regioselectivity) and the desired three-dimensional spatial relations must be obtained (stereoselectivity). The degree to which this type of control can be achieved is shown in the synthesis of the substance adamantane ($C_{10}H_{14}$). This novel molecule resembles in structure a 10-atom "chip" off of a diamond crystal. In a *tour de force,* it was originally produced by a many-step process in only 2.4% yield. Recent research in polycyclic hydrocarbon synthesis now gives adamantane in one step in 75% yield. Then, a surprise practical payoff came when it was discovered that adding a single amine substituent to adamantane gave adamantine (1-amino-adamantane), which proved to be an antiviral agent, a prophylactic drug for influenza, and a drug to combat Parkinson's disease.

The recent development of new ways to construct cyclic five-membered rings has again paid off handsomely. The diverse array of applications includes production of thienamycin, $C_{13}H_{14}N_2SO_4$, a novel relative of penicillin and an important new drug. At another extreme, large ring compounds have been exceptionally difficult to synthesize. Their structures are complicated by functionally crucial left/right-handed structural geometrics ("chiral centers"). Their wide ranging biological properties—from pleasant fragrances for perfumes to anti-fungal, anti-tumor, and antibiotic activities—make large ring synthesis an interesting challenge. An example is erythromycin, $C_{37}H_{68}O_{12}N$, which in the desired atomic hookup can be shaped into 262,144 different structures derived from the many possible ways to couple the right- and left-handedness at its 18 chiral centers (2^{18}= 262,144). Twenty-five years ago, this compound was judged to be "hopelessly complex" by R. B. Woodward, who won the Nobel Prize for synthesizing molecules as complex as quinine and vitamin B_{12}. Today we can aspire to such a goal, in part because of the development of specially designed templates that bring together the terminal atoms of a fourteen-atom chain to form a fourteen-membered ring. This provides the structural framework of erythromycin and it has already resulted in the synthesis of a number of constituents of musk and contributed to our understanding of smell.

■ *Crossing inorganic/organic boundaries.* The traditional line of demarkation between organic and inorganic chemists has virtually disappeared as the list of fascinating metal-organic compounds continues to grow. The ubiquitous appearance of these compounds in biological systems underscores the importance of encouraging this cross-boundary research. Furthermore, research in developing new inorganic substances has provided a surprising dividend in their frequent applicability in organic synthesis.

The latter situation is illustrated by the borohydrides. The cohesive picture we have, at last, for this strange boron/hydrogen family was not possible before their study by X-ray crystallography, infrared and NMR spectroscopy, and molecular orbital theory. Now borohydrides are widely used as selective, mild reducing agents in organic synthesis. Silicon and transition-metal organometallic compounds give other examples. Silicon compounds, for example, are used to fold an extended molecular reactant precisely as needed to synthesize the molecule cortisone. Now this valuable therapeutic agent can be made in less than 20 steps at a yield 1000 times higher than achieved in the earlier 50-step process.

Compounds with metal atoms sandwiched between organic rings continue to proliferate and their chemistry will likely relate to catalysis. The metal-organic complexes provide another active synthetic field, with interest both in catalysis and, perhaps, in radioactive or metal-waste cleanup.

■ *Novel solids.* Chemists are learning how to prepare solids with quite remarkable properties. Thus, families of solid substances are being synthesized which, like the alumina-silicates (the "zeolites"), are deliberately contrived to include particular shaped cavities and channels. Guest molecules that slip comfortably into these channels can be held in favored conformations as reactants bring about desired chemical changes. The result is that only particular molecules (the ones that fit) react and they do so in a structurally specific way. Entirely different sorts of solids are giving new families of electrically conducting materials. The lead molybdenum sulfide, $PbMo_6S_8$, is a superconductor that can remain superconducting up to 600 kilogauss fields. One of the several conducting organic materials is polyacetylene which, on exposure to various chemicals (iodine, arsenic trifluoride, etc.), can approach metallic conductivities. Polysulfurnitride, $(SN)_x$, not only rivals metal conductivity, it also displays superconductivity. Another promising class of new linear-chain polymer conductors are those based upon metallo-macrocycle complexes held together by bridging groups and treated with an electron-accepting compound. This class is amenable to a wide range of electronically significant chemical modification, hence tailoring to intended application.

Pressure as a controlling dimension of solid state reactions has long been recognized but little understood. With techniques now available, solids can be compressed and studied spectroscopically up to a few hundred thousand atmospheres pressure. New molecular structures, reactivities and electronic properties can be obtained. Photochromic materials provide an example. Substances like the anils, spiropyranes, and bianthrones change color when exposed to light. Under pressure, the colors can be "tuned" and some of these materials will display the color change merely by heating, without the need for exposure to light. These effects are attributable to changes in molecular structures and electronic makeup which, in turn, can affect chemical reactivity.

Solid state chemistry is relatively less active in the U.S. than abroad; French, German, and Japanese chemists have leadership positions. Hence, there can be a high return if more activity can be stimulated here in the synthesis and characterization of radically new solids. There will emerge new semiconductors, solid state ionic materials (used in batteries, memory devices, display devices, and chemical sensors), ferroelectrics, pyroelectrics, and magnetic materials. Creative synthetic chemists must be attracted and encouraged in this area.

■ *Pathways using light as a reagent.* Another promising chemical pathway is connected with the use of light photons in chemical synthesis. Many organic molecules display quite different chemistry and structures after absorption of light. Photosynthesis provides the most striking prototype and our understanding is advancing rapidly.

"Artificial photosynthesis" both mimicking Nature and generalizing into new directions is under active study. Chemical storage of solar energy is an obvious long-term goal. In addition, new routes in chemical synthesis are offered. Some high energy ("strained") molecular structures, including those of many natural products, must be formed in energy-consuming (endothermic) processes (e.g., the mycin antibiotics, certain alkaloids, vitamin D precursors, and certain steroid hormones). It is difficult to put in this energy because the active reagents tend to threaten the fragile product. In the photochemically induced process, less aggressive reactants can be used which can be excited with light to approach the product gently from above, with delicate control through the photolysis wavelength. A single example shows the great potential. A calcium homeostatic steroid hormone with the formula $C_{26}H_{44}O_2$ (1,25-dihydrovitamin D_3) can be synthesized photochemically. By using tuned laser light controlled to the optimum wavelength (300nm), the synthetic yield of the desired structure has been increased 400%. How important such light-assisted chemistry will be remains to be seen. Chemists understand the ground state (unexcited) chemistry of literally millions of compounds. Virtually every one of these compounds has different chemistry after light absorption but we have explored only a miniscule fraction of this new domain!

4. Collaboration

The thrust toward *understanding chemical reactivity* has inviting avenues for collaboration across traditional disciplinary boundaries. Much of the most fundamental work on molecular dynamics is at the chemistry-physics interface and it specially benefits from close interaction between chemists and physicists on the one hand and between experimentalists and theoreticians on the other. In a similar way, exciting new horizons in synthesis are bringing together the perspectives and techniques of individuals who used to be differentiated as inorganic or organic chemists. Collaborative mechanisms should be encouraged.

B. Chemical Catalysis

1. Program.

We propose an initiative to apply the techniques of chemistry to obtain a molecular-level understanding of catalysts. Ultimately, new catalyst systems will result that will lay the foundation for the development of new chemical technologies. The program will emphasize research in four areas:

- *Heterogeneous catalysis:* to apply the powerful new instrumental techniques of surface science to the study of chemistry on the surface of a solid.
- *Homogeneous catalysis:* to take advantage of recent developments in synthetic chemistry that show promise as new soluble catalysts.
- *Photocatalysis and electrocatalysis:* to investigate the rich possibilities of solution chemistry catalytically assisted by electrode processes, with and without absorption of light.
- *Artificial-enzyme catalysis:* to bring together chemists' ability to synthesize molecules of predesigned topography and the biochemists' emerging understandings of natural catalysts, the enzymes. We should be able to generate a new class of artificial enzyme-like catalysts tailored to specific needs.

2. *Objectives.*

- To identify the fundamental molecular ingredients of catalytic processes in order, ultimately, to guide the development of chemical high technology.
- To synthesize new catalysts that may serve as the foundation of new chemical processes.
- To increase the number of young scientists who have the breadth and background needed to contribute to the field of chemical catalysis.
- To provide the fundamental knowledge and creative manpower required for the U.S. to maintain competitive advantage in catalysis-aided technologies.

3. *Rationale.*

A catalyst accelerates chemical reactions toward equilibrium without being consumed. This acceleration can be as much as ten orders of magnitude. A *selective* catalyst can have this same dramatic effect but on only one of many competing reactions. While the development of new catalysts was empirical fifteen years ago, research innovations in chemical sciences over the last ten years are converting catalysis from art to science. A familiar example displays the benefits to be gained—the catalytic converter developed for automobiles to reduce air pollution. The catalyst—containing two grams of platinum, palladium, and rhodium per car—oxidizes unburned hydrocarbons and carbon monoxide to harmless water and carbon dioxide. Simultaneously, it reduces toxic nitrogen oxides to harmless nitrogen gas.

Estimates that 20 percent of the gross national product is generated through the use of catalysis demonstrate the major role industrial catalysis has in satisfying such diverse societal needs as food production, energy conversion, defense technologies, environmental protection and health care. On the horizon is the extensive use of catalysts to tap new energy sources—such as heavy oils, coal, oil shale, tar sands, lignites, and bio-mass—as we run out of oil. However, American preeminence in catalysis science is now challenged by Japan, Germany, the USSR, and France. These countries have established institutional networks to accelerate research and development in this evolving high-technology field. To meet this challenge, we must keep our own catalytic science vital.

Industrial research in catalysis is critically dependent on the techniques and concepts which have been developed in university and national laboratories. Since industrial laboratories use the latest surface science and laser spectroscopic techniques, they want their new young scientists to bring with them state-of-the-art experience. Furthermore, industrial catalysis research must be chosen with an eye toward existing technology and product lines. Industry cannot sustain the adventurous fundamental research that will underlie the discovery of radically new catalysts. To illustrate this, consider the enormously important conversion of atmospheric nitrogen into ammonia. The original work by Haber in Germany produced an ammonia yield of 13% per pass. Today the same process is used on a massive scale. Yet, seventy years after Haber's work, a modern ammonia unit producing some 1500 tons per day operates at only slightly greater yield per pass (perhaps 15%). There is no physical limitation here—equilibrium is on our side. What is needed is fundamental research in catalysis. This is a role for university chemists. It is a place for federal investment.

This chemical catalysis program will concentrate on long-term research to learn about catalytic processes on the molecular level. Then the new fundamental knowledge generated will be transferred into technology by the employment of the emerging young scientists in industry and by the ties formed in this program among scientists in the universities, the national laboratories and industry.

HETEROGENEOUS CATALYSIS

Heterogeneous catalysts are solid materials prepared with large surface areas (1-500 m^2/g) upon which chemical reactions occur at extremely high rate and selectivity. Some major new commercial processes based on heterogeneous catalyst developments in recent years include:

—"reforming" hydrocarbons to high octane composition (platinum alloy catalysts);
—"cracking" high paraffins to gasoline (molecular sieve catalysts);
—methanol to gasoline, jet fuel (molecular sieve catalysts);
—exhaust converter on automobiles (platinum, palladium, rhodium catalysts);
—ethylene oxidation to ethylene oxide (silver, cesium, chlorine catalyst);
—propylene oxidation to acrolein and acrylonitrile (bismuth, and molybdenum oxide catalyst);
—ethylene polymerization (chromium catalyst);
—propylene polymerization (titanium, magnesium oxide catalyst).

Surface science is developing rapidly and now gives us experimental access to this two-dimensional reaction domain. Because of the unsatisfied bonding capability of the atoms at the surface, chemistry here is dramatically different from that of the same reactants brought together in solution or the gas phase. But when chemists are able to "see" what molecular structures are on the surface, then all of our knowledge of reactions in conventional settings becomes applicable. This will open the door to understanding and controlling the chemistry in this surface domain. There are five areas of heterogeneous catalysis where this understanding will have major impact on new chemical technologies.

■ *Molecular sieve synthesis and catalysis.* Molecular sieves are crystalline alumina-silicates containing pores or channels within which chemical reactions can be initiated. They offer unparalleled efficiency both for cracking of petroleum and for conversions such as shale oil or methanol to gasoline. We need to know better how to synthesize molecular sieves with controlled molecular pore size, as well as to determine the elementary reaction steps and intermediates that account for their efficacy.

■ *Metal catalysis.* Finely dispersed transition metals are already coming into use to catalyze hydrocarbon conversions and ammonia synthesis for fertilizers. Other such applications and improved performance will follow from intensive research into the control of surface structures, oxidation states, residence time of reaction intermediates, and resistance to catalyst "poisons" (such as sulfur).

■ *Substitutes for precious metal catalysts.* Many of the most effective catalysts are rare metals not available in the U.S., including rhodium, platinum, palladium, and ruthenium. Their strategic value requires a concerted research effort to find more accessible substitutes, such as transition metal oxides, carbides, sulfides, and nitrides.

- *Conversion catalysts.* We must find catalysts to convert abundant substances to useful fuels and industrial feedstocks. These reactions include conversion of atmospheric nitrogen to nitrates, methane to methanol, carbon dioxide to formate, and depolymerization of coal and biomass to useful hydrocarbons.
- *Catalysts to improve air and water quality.* We have many environmental pollution problems for which we need to match the spectacular success of the catalytic converter used to clean automobile exhaust gases. To begin, we need catalysts that remove sulfur oxides from smoke plumes, that purify water, and that prevent acid rain.

As we learn more about the molecular structures at the solid-gas interface (reactants, intermediates and products), a better understanding of surface chemical bonding will follow. We can look forward to understanding additives that modify catalyst performance ("promoters" and "poisons"). Then, the challenging subject of synthesis of the designed catalyst can be addressed. All of this fundamental knowledge will underlie and facilitate the development of new and more selective heterogeneous catalysts.

HOMOGENEOUS CATALYSIS

Homogeneous catalysts are soluble and active in a liquid reaction medium. Often they are complex metal-containing molecules whose structures can be modified to tune reactivity in desired directions to achieve very high selectives. (In this sense homogeneous catalysts can be superior to heterogeneous ones.) The largest industrial-scale process using homogeneous catalysis is the partial oxidation of para-xylene to terephthalic acid (United States production in 1981 was 6.2 billion pounds per year), which uses dissolved salts of cobalt and manganese as the catalyst system. Most of the product ends up polymerized to give us polyester clothing, tire cord, soft drink bottles, and a host of other useful articles.

An important branch of homogeneous catalysis has developed from research in organometallic chemistry. An example is rhodium dicarbonyl diioxide, employed in the commercial production of acetic acid from methanol and carbon monoxide. With this catalyst present, the reaction economically gives more than 99% selectivity to acetic acid.

There are three areas of homogeneous catalysis with potential for major impact on new chemical technologies from increased understanding.

■ *Activation of inert molecules.* There are a number of relatively inert molecules that are enticing as reaction feedstocks because of their abundance: nitrogen, carbon monoxide, carbon dioxide, and methane. One way to approach this end might be through homogeneous organometallic catalysis. Dramatic examples of this promise are beginning to appear. Soluble compounds of tungsten and molybdenum with molecular nitrogen have been prepared and induced to produce ammonia under mild conditions. The carbon-hydrogen bonds in normally unreactive hydrocarbons have been split by organorhodium and organoiridium complexes. Hope for buildup of complex molecules from one-carbon molecules, such as carbon monoxide, is encouraged by recent demonstrations of carbon-carbon bond formation at metal centers bound in soluble metal-organic molecules. Synthesis of compounds with multiple bonds between carbon and metal atoms has had a major impact in clarifying the catalytic interconversion of olefins. While there is much to learn, the stakes are high and the odds for success are excellent.

■ *Metal cluster chemistry.* An adventurous frontier of catalysis lies in the expanding capability of chemists to synthesize molecules built around several metal atoms in proximity (a "metal cluster"). In parallel, solution and cryogenic techniques are revealing the structures and chemistry of small aggregates containing only metal ions or atoms ("naked clusters"). All of these clusters, bound or "naked," furnish a natural bridge between homogeneous catalysis and bulk metal, heterogeneous catalysis.

Many directions are being explored. Cubical units of four metal atoms and four sulfur atoms are now known for iron, nickel, tungsten, zinc, cobalt, manganese, and chromium. This "cubane" structure for iron has been found to be the functional unit in the ferrodoxin iron proteins that catalyze electron transfer reactions in biological systems. Many cluster compounds have been made from metals bound to carbon monoxide. These metal carbonyls have formulas $M_x(CO)_y$, and x can be made very large. (The world's record as of this writing is a platinum compound with $x=38$.) It is intriguing that many of the most catalytically active metals also form cluster compounds (e.g., rhodium, platinum, osmium, ruthenium, iridium, etc.). Now the chemistry of these elements can be studied as a function of cluster size.

■ *Stereoselective catalysts.* Another exciting frontier involves the development of homogeneous stereoselective catalysts. Many biological molecules can have either of two geometric structures connected by mirror-image (chiral) relationships, and generally only one of these structures is functionally useful in the biological system. If a complex molecule has seven such chiral carbon atoms and a synthetic process produces all of the mirror-image structures in equal amounts, there would be $2^7 = 128$ structures, 127 of which might have no activity or, worse, some undesired effect. Thus the ability to synthesize preferentially the desired structure with the desired geometry at every chiral center is essential.

Immense strides in this area are being made. For example, L-dopa, an amino acid that has revolutionized the treatment of Parkinson's disease, is now made using an asymmetric addition of hydrogen to a carbon-carbon double bond. The catalyst is a soluble rhodium phosphine catalyst that gives 96% of the correct structure with high efficiency. Stereospecific oxidations can also be carried out. The recent invention of a titanium catalyst to add, in a specific geometry, an oxygen atom across a carbon-carbon double bond has lowered the price of gypsy moth attractant ten-fold. Despite these successes, the basic factors that produce stereochemical control are not at all well understood. Mechanistic studies are needed and the rewards from better understanding will be great.

Future advances in homogeneous catalysis are dependent upon easy access to advanced instrumental techniques: X-rays, high-field nuclear magnetic resonance, electron spin resonance, mass spectroscopy, and computational facilities. Further, work at the interfaces between organic, inorganic, and physical chemistry is involved, so breadth of knowledge is especially important.

PHOTOCATALYSIS AND ELECTROCATALYSIS

Exciting advances have recently been made in the study and control of chemistry taking place at the interface between a liquid solution and an electrochemical electrode surface. In some applications, the chemistry is initiated by absorption of light by a semiconductor used as an electrode. Whether light is involved (photocatalysis) or not (electrocatalysis), this rapidly moving field depends upon our knowledge both of homogeneous catalysis and of semiconductor behavior.

■ *Photoelectrochemical cells.* There is considerable promise from research directed toward conversion of light to electrical energy in a photoelectrochemical cell. In such a cell, one or both of the electrodes is a semiconductor material that absorbs the incident light. The lure is that semiconductors, probably with chemically modified surfaces, can absorb red and infrared light, the spectral region within which most solar energy falls.

Chemists are learning how to bond dyes and protective polymers to the surface of a semiconductor solid through covalent bonds. There is now the potential for solving the key problems of photoelectrochemical cells: to catalyze the essential electron-transfer reactions with solution species, to prevent chemical erosion of the electrode, to avoid back-reaction, and to shape the active spectral region to match that of the energy source, the sun. Already cells for conversion of solar energy to electrical energy have rivaled those of solid state photovoltaic cells (e.g., the n-GaAs/selenide system operates at 14% efficiency). Numerous "non-conventional" semiconductors are under investigation, such as TiO_2, CdS, WSe_2. These materials are still speciality products but there is good prospect that they will cost much less than existing single crystal photovoltaic materials.

■ *Photocatalysis.* Related studies involving photoelectrochemical cells or suspensions of semiconductor materials focus attention on the chemistry that can be brought about. In these systems, light absorbed in the semiconductor promotes catalytic oxidation-reduction chemistry at the electrode-solution or at a membrane-solution interface. Such oxidation-reduction chemistry has significant scientific interest and, likely, great practical significance. For example, photodestruction of a toxic waste material such as cyanide has been demonstrated at titanium dioxide surfaces. A more popularized and perhaps feasible concept is that such photo-catalytic chemistry, solar energy driven, could produce massive amounts of hydrogen and oxygen from the photoelectrolysis of water. What an intriguing prospect, to convert from diminishing and seriously polluting petroleum fuels to a renewable fuel that burns to water and that is made from water using energy from the sun.

■ *Electrocatalysis.* Apart from light-initiated processes, electrode surfaces with catalytic activity offer a new domain for chemical synthesis. In a field with a long heritage, recent developments have shown that electrode surfaces can be chemically tailored to promote particular reactions. This research area has adopted techniques from the semiconductor industry (such as chemical vapor deposition) and coupled them with imaginative synthetic chemical techniques for surface modification.

An example is the electrocatalyst family developed for use in chlorine generation in chlor-alkali cells. A successful case is based upon a thin layer of ruthenium dioxide (the catalyst) deposited on a base metal electrode. This electrocatalyst has dramatically changed practice in the chlor-alkali industry (an industry representing billions of dollars in sales) because of its improved energy efficiency and reduced cell maintenance. Future developments will include radically improved fuel cells, which provide clean and thermodynamically efficient conversion of chemical fuels to electricity.

■ *Chemistry at the solid/liquid interface.* Before the technological potentialities of any of the above can be realized, we must have a much better understanding of the nature of chemistry at the semiconductor/liquid interface. The marvelous instrumentation developed for surface science studies is, of course, applicable only at solid/vacuum interfaces. We need comparable capability at the solid/liquid boundary. This capability will be won from fundamental research in solid state chemistry, electrochemistry, surface analysis, and surface spectroscopy. Where the biggest gains are to be made is a speculative question. The surprising discovery of the million-fold intensification involved in surface-enhanced Raman effect encourages optimism.

The potential gains from these exciting areas are considerable. We need to know how to catalyze multielectron transfer reactions at an electrode surface. That is the chemistry required, for example, to photogenerate a liquid fuel such as methanol from carbon dioxide and water. Multielectron transfer catalytic electrodes for oxygen reduction in electrochemical cells would find a welcoming home in the fuel cell industry.

It is also likely that research on semiconductor electrode surface modification will reflect back beneficially into the field of electronics. Thus, the integrated circuit technology based upon GaAs may depend upon control of its surface chemistry. Already photoresist/chemical etching techniques are recognizing the importance of the chemistry involved in surface modification (witness the pursuit of "anisotropic chemical etching").

In summary, our evolving understanding of the electrode/solution interface, buttressed by concepts based on semiconductor electrodes and the development of a number of new methods for modifying electrode surfaces, has provided powerful new approaches to both photocatalysis and electrocatalysis. Future advances will benefit synergistically from progress in heterogeneous and homogeneous catalysis, increased understanding of mass and charge transport within the electrode surface layers, and continued development of experimental methods and theoretical models for the interface.

ARTIFICIAL-ENZYME CATALYSIS

The most striking benefit from our bounteous knowledge of reaction pathways and the analytical capacity of modern instrumentation has been the development of capability to deal with molecular systems of extreme complexity. With the synthetic chemist's prowess and such diagnostic instruments as nuclear magnetic resonance, X-ray spectroscopy, and mass spectroscopy we can now synthesize and control the structure of molecules that approach biological complexity. This includes the ability to fix the molecular shape, even extending to the mirror-image properties that are so crucial to biological function.

There is no application of these capabilities more intriguing than coupling them with our growing knowledge of catalysis to synthesize artificial enzymes. In Nature, enzymes are the biological catalysts that accelerate a wide variety of chemical reactions at the modest temperatures at which living organisms can survive. A given enzyme selects from a many-component system a single reactant molecule and transforms it to a single product with prescribed chiral geometry.

Without catalysts, many simple reactions are extremely slow under ambient conditions. Raising the temperature speeds things up, but this may include a variety of undesired outcomes—acceleration of unwanted reactions, destruction of delicate products, and waste of energy. Hence, there are strong reasons to develop synthetic catalysts that work like enzymes. First, natural enzymes do not exist for most of the chemical reactions in which we have interest. In the manufacture of polymers, synthetic fibers, medicinal compounds, and many industrial chemicals, very few of the reactions used could be catalyzed by naturally occurring enzymes. Even where there are natural enzymes, their properties are not ideal for chemical manufacture. Enzymes are proteins, sensitive substances that are easily denatured and destroyed. In those industries that do use enzymes, major effort is directed toward modifying them to make them more stable.

■ *Controlled molecular topography and designed catalysts.* We have a pretty good idea how enzymes work. Nature contrives a molecular surface fitted to a specific reactant. This surface attracts from a mixture the unique molecular type desired and immobilizes it. Held on the surface, the selected reactant is forced to abandon its flexibility and assume a distinct shape. When the reaction partner arrives, the scene is set for the desired reaction to take place in the desired geometry.

Organic chemists who have taken up this challenge are making excellent progress. Without special control, large molecules usually have exclusively convex surfaces (ball-like shapes). So a first step has been to learn to synthesize large molecules with concave surfaces. Then, these concave surfaces could be shaped to accommodate a desired reactant. Cyclodextrins provide examples—they are toroidal in shape. The crown ethers, developed over the last 15 years, have a quite different surface topography. For instance, 18-crown-6 consists of twelve carbon atoms and six oxygen atoms evenly spaced in a cyclic arrangement. In the presence of potassium ions, the ether takes up a crown-like structure in which the six oxygen atoms point toward and bind a potassium ion. Lithium and sodium ions are too small and rubidium ions too large to fit in the crown-shaped cavity, so this ether preferentially extracts potassium ions from a mixture. Much more ornate examples now exist. Chiral binaphthyl units can be coupled into cylindrical or egg-shaped cavities. With benzene rings, enforced cavities have been made with the shapes of bowls, pots, saucers, and vases.

Plainly we are moving toward the next step, to build into these shaped cavities a catalytic binding site. The earliest successes will likely be patterned after natural enzymes but there is no doubt that in time, artificial, enzyme-like catalysts will not be limited by what we can find already known in Nature.

■ *Biomimetic enzymes.* A short-cut approach is to pattern artificial enzymes closely after natural enzymes. This has been called "biomimetic chemistry." As examples, mimics have been prepared for the enzymes that biologically synthesize amino acids. The artificial enzymes incorporate the important catalytic groups of a natural enzyme (vitamin B_6, for example) and show good selectivity, even including the formation of the correct chirality in the product. Mimics have been prepared for several of the common enzymes involved in the digestion of proteins, e.g., substances that catalyze the cleavage of RNA have been synthesized based upon the catalytic groups found in the enzyme ribonuclease. Mimics have also been synthesized which imitate the class of enzymes called cytochromes P-450, which are involved in many biological oxidations, and the oxygen carrier hemoglobin. Furthermore, mimics have been prepared for biological membranes and for those molecules which carry substances through membranes. These have potential applications in the construction of organized systems to perform selective absorption and detection, as in living cells.

It is important for the U.S. to build on its early lead in this field. Although most of the work mentioned above has been done in the U.S., the Japanese have also become extremely active and have specifically targeted "biomimetic chemistry" as an area of special opportunity. Research on synthetic organic chemistry develops novel methods to construct the required molecule, and elaborates new kinds of structures. The study of detailed reaction mechanisms in organic and biological chemistry permits a rational approach to catalyst design. This is an area ripe for development, deserving special encouragement as a part of this program in Chemical Catalysis.

4. Collaboration.

Thus, catalysis research with the four facets named here involves many subfields of chemistry (surface chemistry, solid state chemistry, organic synthesis, inorganic synthesis, photochemistry, electrochemistry, bioorganic chemistry and chemical engineering) along with the contiguous fields of solid state physics, surface physics, biology, and biochemistry. Each of these different perspectives is needed to address fully the common goal, to understand how a catalyst intermediary can enhance preferentially the rate at which a desired reaction takes place. Hence, there is great opportunity for collaboration across disciplinary boundaries.

C. Chemistry of Life Processes

1. Program.

We propose an initiative to develop and apply the techniques of chemistry and the skills of individuals knowledgeable in that discipline to the solution of molecular-level problems in life processes. This initiative includes both the application of chemical science to biology, and the development of new chemical science based on stimuli provided by biology. Research at this border between chemistry and biology requires individuals broadly competent in both areas, and a special effort must be made to develop such individuals.

The program will emphasize the molecular aspects of six areas.

- Enzymology: to understand the molecular interactions responsible for natural enzymatic activity and its inhibition; to produce natural enzymes for use in chemical synthesis.
- Immunochemistry: to learn the chemical basis of the immune and allergenic response; use, function, and modification of monoclonal antibodies; synthesis of antigens and adjuvants.
- Chemical Endocrinology: to synthesize hormones and hormone analogs; to understand hormone agonist and antagonist mechanisms.
- Neurochemistry: to determine the molecular basis of nerve transmission, neurotransmitters, agonist and antagonist chemistry, and membrane polarization.
- Membrane Chemistry and Vectorial Chemistry: to clarify active and passive transport (bioenergetics).
- Biological Model Studies: to understand host/guest chemistry, semisynthetic enzymes, and active sites.

2. *Objectives.*

- To strengthen basic research in the life sciences by providing the molecular-scale analytical, synthetic, and structural techniques required to identify, prepare, and study the organic chemical compounds which are central to modern molecular biology.
- To accelerate the conversion of qualitative biological information into techniques and substances useful in human and animal medicine and agriculture.
- To educate graduate students and postdoctoral fellows in both chemistry and biology; to provide individuals skilled in chemistry (especially organic synthesis, organic reaction mechanisms, and analytical chemistry), who can understand and interpret biological problems and opportunities in molecular terms.

3. *Rationale.*

The study of the molecular bases of the reactions occurring in life offers enormous opportunity both for discovery in basic physical science and for application to human health, animal health, and agriculture. This area of research includes some of the most important problems which now exist in science. Solutions to these problems will have very large and beneficial impacts on society.

Biology has evolved a range of problems of revolutionary significance that now require analysis in terms of molecular interactions. The molecular synthesis and analysis required for the solutions of these problems can only be provided by chemistry.

A major fraction of biology can be reduced to two broad types of problems at the molecular level:

RECEPTOR-SUBSTRATE INTERACTIONS

Essentially all biological processes are mediated by the selective interaction of a protein receptor (enzyme, antibody, membrane and intracellular receptors) with one or more specific substrates (enzyme substrate, antigen, hormone, neurotransmitter, or simple molecule or ion).

To study these processes in molecular detail, and to control them with the precision required for rational applications in medicine, agriculture, and other areas, it is necessary to be able to:

- isolate and identify the structures of endogenous substrates or substrate analogs;
- synthesize these substrates and analogs in useful quantities and high purity;
- analyze their interactions with their receptors both in physical-chemical and biological terms;
- modify their structures to give useful agonist and antagonist activity and to facilitate effective delivery into living organisms.

The development of analytical, structural, synthetic, and mechanistic techniques appropriate to this area of molecular science represents an enormous challenge and opportunity for chemistry. The active substances may be present *in vivo* only in minute quantities; they are usually water soluble (polypeptides, polysaccharides) or surface active (lipids, membrane or hydrophobic proteins); they may have high molecular weights and complex or heterogeneous structures. Medicine and biology need the methods of chemistry to manipulate these substances; chemistry cannot presently provide fully adequate methods. Hence, the challenge and the opportunity for chemists.

MEMBRANES AND VECTORIAL CHEMISTRY

The set of chemical reactions that constitute life takes place in cells and organelles. The concept of vectorial chemistry (reactions depending upon spatial separation of reactants into regions of different concentrations—that is, on concentration gradients—for their driving force) is poorly understood at a molecular level, but essential to biochemistry (the Mitchell chemiosmotic hypothesis is the best-known example).

Understanding the relations between concentration gradients across membranes and processes which form or break chemical bonds is a formidable task, one which requires active chemical modeling at both the synthetic and the kinetic level to complement the current activity in mechanistic biology. The related problem of the transport of molecules across membranes (for example, between blood and a particular type of cell) is central to the area of pharmacokinetics and drug delivery.

What reason is there to expect chemistry to be able to make progress on the many specific problems contained in these two major general classes of problems? A number of recent advances in science and engineering have provided an extraordinarily powerful set of techniques which are appropriate for the investigation of problems in chemical biology at the molecular and supramolecular level. Examples are given below.

- Recombinant DNA technology provides the opportunity to prepare proteins in substantial quantities with controlled variations in amino acid sequence.
- Monoclonal antibodies produced in large quantities using hybridoma techniques provide specific receptors against virtually any high or medium molecular weight substance.
- X-ray single crystal analysis can now solve very complex structural problems.
- Electron microscopy permits direct visualization of many structures at levels between 2 and 1000 A and is superbly suited to solving problems in membrane chemistry and in the structures of large molecules.
- Ultrasensitive analytical techniques (fast-atom-bombardment mass spectrometry, two-dimensional NMR spectroscopy, antibody-based competitive binding assays, electron capture gas-liquid chromatography, high performance liquid chromatography) can elucidate structures of small quantities of very complex substances.
- Sophisticated synthetic technology, based both on conventional chemical techniques and on a range of biological techniques, including enzymatic transformation and fermentation, make it possible to synthesize complex naturally occurring substances and analogs in useful quantities.

- Computers permit analysis of the details of receptor-substrate binding, and modeling of the complex set of coupled reactions which take place in living cells.

In summary, molecular biology has become sufficiently detailed that many of its most important current questions must be phrased and answered using chemical methods. Hence, chemical techniques are essential to understanding biological phenomena and for converting this understanding into a form useful in medicine and agriculture. There is a pressing practical reason for accelerating this application of chemistry to biological problems, namely, the economic and technological relevance of chemical biology to biotechnology and applied biology, and the intense international competition in this area. Failure to encourage the chemical aspect of biology will cost us leadership in this area and we will forfeit ownership of the practical results to other countries.

4. *Collaboration.*

An important aspect of the proposed program would be its emphasis on interdisciplinary research. Where appropriate, the program would actively encourage development of collaborative projects which integrate programs of chemical and biological research.

In addition, biotechnology is an area of extremely active industrial research. This derives from the special potential for health care (with efforts by essentially all pharmaceutical companies and by a host of small start-up biotechnology companies) and improvements in agriculture (Monsanto, du Pont, and many others). The major immediate markets being addressed by these companies are high-value proteins (interferon, plasma thrombinogen activator, insulin, serum Factor VIII, animal and human growth hormones, vaccines) and monoclonal antibody products (clinical diagnostics). Longer-term industrial projects are aimed at a range of more complex health-care products (treatments for cancer, diabetes, high blood pressure; methods for organ transplantation; immuno-regulators), improved plant strains, enzymatic catalysts for industrial and food use, biomass utilization for energy, and human genetic engineering.

The ultimate success of this industrial endeavor will depend heavily upon the supply of young scientists with the needed cross-disciplinary expertise. This program will provide such experience through research more basic than but clearly complementary to the major threads of current industrial research. For example, the development of molecular understandings of the reactions in nerve transmission and of oxidative phosphorylation are central to advances in the relevant biology but still far from application. Ultimately, these areas will be central to the development of neuroactive drugs and more efficient plants. Similarly, the development of new synthetic methods for biologically relevant molecules remains primarily an academic activity, but it is critical to the continued leadership of our pharmaceutical industry.

In fact, the scope for new research in chemical biology is so great that full and cooperative effort from both university and industry is both warranted and required. Industry is now actively building on the basic science of recent years (as well as conducting some of its own very excellent basic work); the university must provide the basic science for the next cycle of industrial discovery and development.

III. CONCLUSIONS

Chemistry extends all the way from physics, with its strong ethos of deductive logic, to the largely phenomenological biological sciences. We are at a gratifying period when chemistry is "hot at both ends." With adequate access to the powerful techniques now developing, chemistry can expect to explore and understand the principles that govern chemical change at a much more fundamental level than has been possible in the past. At this end of the spectrum, chemistry can begin to escape the bounds of empiricism imposed thus far by the intractability of many-body theory and the limits of our measurements. Yet, despite these barriers, chemists have been so enormously successful with their inductive approach that molecules of exquisite complexity can be recognized and synthesized in every intricate detail. This capability is most timely, because the phenomenological advances of the biosciences now demand explication at the molecular level. Thus, the time is ripe for quite spectacular advances in chemistry.

Hence, in the next ten to twenty years, there will be dramatic changes in our basic understandings of chemical change and our ability to marshal those understandings to deliberate purpose. The program presented here is intended to define a leadership role for the U.S., as these advances are won. The rewards accompanying such leadership are commensurate with the prominent role of chemistry in addressing societal needs, in ameliorating the problems of our technological age, and in sustaining our economic well-being. The costs of falling behind are simply intolerable.

To exploit the intellectual opportunities in chemistry and to maximize the social gain, a substantial incremental increase in federal support for basic research in the chemical sciences is needed. This need for enhanced federal investment in chemical research is rooted in a pattern of funding historically appropriate to a test-tube and bunsen burner era, an era long since eclipsed. The sophistication of a modern chemistry laboratory requires a much more vigorous and sustained financial commitment, both in capital investment and in supporting services. The cost is miniscule compared to the stakes involved; we must nurture a $175 billion industry that maintains a $12 billion positive balance of trade and employs over a million people. We must continue to provide it with a full reservoir of fundamental knowledge.

APPENDIX I
FUNDAMENTAL QUESTIONS BEFORE US

A. Understanding Chemical Reactivity

Molecular Dynamics.

a. How is energy redistributed as a function of time if it is deposited in a molecule in a specific excitation?
b. How does the energy in a molecule (amount, states, intramolecular distribution) determine its reactivity?
c. Can the chemistry of a molecule be affected by an intense (but not absorbed) photon field?
d. Can we understand and theoretically predict the detailed energy distribution among the products in bimolecular reactions?
e. Can we observe the transition state in real time?
f. Is there a Franck-Condon aspect to H or H + transfer?
g. How can a non-thermalized distribution of excited states be maintained and can novel chemistry be achieved thereby?
h. Can we fully characterize and understand intramolecular and intermolecular energy transfer processes?
i. Can we store solar energy using artificial photosynthesis?
j. What new reaction mechanisms operate at high temperatures because of unusual gaseous molecules and unusually rapid mobility?
k. Can we understand and predict the unusual chemical species and solid phases formed at high temperatures?

Chemical Pathways.

a. What are the systematics of the metal-ligand bond energies in organometallic compounds?
b. What are the mechanisms and rates of multi-electron transfer reactions and how can they be controlled?
c. Can interfacial energy be used to synthesize new solids with unique structural features (e.g., channels, cavities, holes)?
d. How can metastable non-stoichiometric solids be prepared?
e. How can the dimensionality of a solid be controlled (linear chains, two dimensional nets, sandwich solids, etc.)?
f. What principles govern conformer-selective chemistry?
g. Can synthetic polymers that are dynamic entities be designed?
h. Can novel inorganic solids be made with molecular beam epitaxy?
i. Do topological principles underlie some aspects of chemical structures and reactions?
j. What molecular properties exhibit bistability suitable for digital information storage in a molecular electronics system?
k. Can organic ferromagnetic compounds be synthesized?
l. What are the structural requirements for nonlinear optical properties in organic molecules?

B. Chemical Catalysis

Heterogeneous Catalysis.

a. For chemisorbed intermediate species, what are the residence times? Can they be changed to raise catalyst efficiency?
b. For chemisorbed intermediate species, what molecular structures and oxidation states give optimum selectivity?
c. Why are transition metals the best catalysts?
d. What general mechanisms determine catalyst deactivation? Specifically, how do nickel and vanadium compounds deactivate catalysts?
e. What is the combusion process of carbon on catalytic surfaces?
f. What determines surface migration of chemisorbed species on a surface? What role does this migration play in catalysis?

g. Can we identify and characterize active sites in heterogeneous catalysts?

Homogeneous Catalysis.

a. How is the rupture of H-H, C-H, and C-C bonds altered as the size of a metal cluster catalyst increases?
b. What relationships connect homogeneous, metal cluster, and metal surface catalysis?
c. Can we design homogeneous catalysts that would selectively convert coal to specific desired products?
d. How can chiral metal complexes be designed to induce chirality in organic reaction products?
e. Can we achieve selective homogeneous oxidation catalysis for terminal positions in alkanes?
f. Can catalytic Diels-Alder chemistry be developed?
g. Can oxidative fixation of nitrogen (e.g., to nitric acid) be catalyzed?
h. Can the addition of water or ammonia to olefins (to give alcohols and amines) be catalyzed?

Photocatalysis and Electrocatalysis.

a. How does catalysis at the solid-gas interface connect to electrocatalysis at the solid-liquid interfaces?
b. Can we define, control, and understand the electronic structure of surfaces?
c. What principles are involved in the design and synthesis of new electrocatalysts?
d. What principles are involved in the design and synthesis of new semiconductors materials for photoassisted reactions?
e. Can we develop analytical methods to study chemistry at the solid/liquid interface to match the power of those recently developed for use at the solid/vacuum interface?
f. How can light-absorbing dyes be chemically bound to semiconductor surfaces?

Artificial-Enzyme Catalysis.

a. What obstacles impede synthesis of artificial-enzyme catalysts tailored for a desired reaction?
b. Can we design catalysts to fix the structure (head-to-tail, head-to-head, block, etc.) and stereochemistry of polymers?
c. Can we enhance selectivity through surface shape?

d. Can artificial enzymes be designed to operate in organic solvents?
e. Can artificial enzymes be incorporated into artificial membranes to achieve energy-driven transport (as occurs in living cells)?
f. By adding binding sites for regulator molecules, can we design artificial enzymes to act as detectors? Can these be developed with feedback loops to adapt and control their environment?

C. Chemistry of Life Processes

Enzymology.

a. What is the relation between receptor protein/enzyme binding/active site structure and specificity/catalytic activity?
b. What strategies are required to design substrates which compete with endogeneous substrate for specific receptor/active sites?
c. Can enzymes be used as practical catalysts in organic synthesis?
d. What are the mechanisms of enzymatic reactions (e.g., in protein and nucleic acid synthesis, oxidative phosphorylation, photosynthesis, active transport, intermediary metabolism)?
e. Can combined chemical/biological synthetic techniques be developed to improve markedly the synthesis of nucleic acid sequences?
f. What is the atomic-level basis for enzyme tertiary structure and stability?

Immunology.

a. What is the relation between immunoglobin structure and binding specificity?
b. What are the best techniques for producing and purifying monoclonal antibodies?
c. What antigens and adjuvants are most effective in initiating an immune response?
d. How large a fragment of a protein antigen is required to evoke an effective immune response?
e. Can monoclonal antibodies serve as a basis for chemical purification procedures? As starting materials for semisyn-

thetic enzyme production? For new types of chemical analytical schemes?

Endocrinology.

a. What are the structures of important endogenous hormones?
b. What is the best strategy for inferring the binding specificity of hormone receptors?
c. How should one design synthetic hormone analogs that have desired activity, specificity, and *in vivo* lifetime?
d. What synthetic strategies are required to synthesize synthetic hormones and analogs (especially polypeptide hormones)?

Neurochemistry.

a. What are the structures of important neurotransmitters?
b. What related structures will have agonist/antagonist activity?

Membrane Chemistry and Vectorial Chemistry.

a. Is the chemiosmotic hypothesis correct? What is the molecular basis for interconversion of energy stored in concentration gradients and energy stored in reactive chemical bonds?
b. How is charge separation across membranes accomplished in photosynthesis, respiration, and related areas of bioenergetics?
c. What is the molecular basis for active transport processes? Can understanding of these processes be used to suggest new types of abiological membranes having useful selectivity?
d. What is the chemistry of perception (vision, smell, memory)?
e. What is the relation between chemical, physical, and biological properties of lipid membranes? Can synthetic membrane systems (liposomes, others) be developed for drug delivery?

Chemical Studies Relevant to Biological Systems.

a. Can we understand the hydrophobic effect in aqueous solutions?
b. What are the best strategies for the synthesis of water-soluble molecules having biological activity (oligosaccha-

rides, lipids, polypeptides, neuroactive substances, nucleic acids, metabolites)?

c. What are the structural and thermodynamic bases for molecular recognition (host/guest chemistry)?

d. What are the properties of thin hydrocarbon membranes (selectivity, permeability, concentration polarization, interfacial structure, thermodynamics of membrane diffusion)?

e. How are cooperative chemical phenomena related to biological systems (liquid crystals, organized lipid assemblies, oscillating reactions, nucleation and phase transitions)?

f. Can one construct instructive chemical models for important biological phenomena (active transport, photosynthesis, oxidation phosphorylation, proton translocation, muscular contraction, information storage, nerve conduction, self-replicating systems)?

THE REPORTS OF THE RESEARCH BRIEFING PANELS

1984

RESEARCH BRIEFING PANEL ON COMPUTER ARCHITECTURE

Jacob T. Schwartz ***(Chairman)***, *New York University, New York, N.Y.*

Richards Adrion, *National Science Foundation, Washington, D.C.*

Donald Bonstrom, *ETA Systems, St. Paul, Minn.*

Charles Bostick, *National Security Agency, Fort Meade, Md.*

John Cavalini, *Department of Energy, Washington, D.C.*

Robert J. Douglass, *Los Alamos National Laboratory, Los Alamos, N. Mex.*

W. Daniel Hillis, *Thinking Machines Corporation, Waltham, Mass.*

Thomas A. Keenan, *National Science Foundation, Washington, D.C.*

David Kuck, Kuck and Associates, *Champaign, Ill.*

H. T. Kung, *Carnegie-Mellon University, Pittsburgh, Pa.*

Thomas Nash, *Fermi National Accelerator Laboratory, Batavia, Ill.*

James Pomerene, *IBM Corporation, Yorktown Heights, N.Y.*

John Riganati, *National Bureau of Standards, Gaithersburg, Md.*

Arnold Reisman, *Microelectronics Center of North Carolina, Research Triangle Park, N.C.*

Paul Schneck, *Office of Naval Research, Arlington, Va.*

Stephen Squires, *DARPA/IPTO, Arlington, Va.*

Kenneth Wallgren, *NASA, Washington, D.C.*

Staff

Mark Weiser, ***Rapporteur,*** *University of Maryland, College Park, Md.*

Gary E. Clark, ***Staff Director,*** *Computer Science and Technology Board*

Rosena A. Ricks, ***Administrative Secretary,*** *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 ARCHITECTURE

I. INTRODUCTION

The market for supercomputers is only a small part of the total computer market. However, it represents a critical cutting edge for the advance of computer technology and provides machines that are critical to technological leadership in a whole range of areas varying from aircraft design to basic chemistry. For this reason it is vital that the programs that achieved U.S. supercomputer leadership be continued and that new programs to maintain that leadership be initiated.

Rapidly developing VLSI technology has created the basis for major advances in superspeed computer performance. Maximum attainable computation rates, which are approaching 1,000 MIPS (millions of instructions per second) for the largest vector supercomputers, will rise over the next decade through 20,000 MIPS, and possibly to levels of 100,000 MIPS and more. Attaining such extreme speeds will depend on the effective use of parallelism, i.e., on the development of computers that can execute many hundreds, thousands, or tens of thousands of instructions simultaneously and of software that can effectively orchestrate many simultaneous streams of computation. The ability to explore this essential new opportunity has been opened up by the amazing reductions in circuit costs attained during the past decade.

This document will review the current supercomputer situation and will predict developments over the next 5 to 10 years. The following major areas will be considered:

1. the *chip and mass storage technologies* on which U.S. primacy in the computer area, and its ability to advance, ultimately rest;
2. the *computer architectures*, both novel and relatively conventional, likely to be important for large-scale computing over the next decade;
3. the potentially significant role of *special-purpose computers*, which may be of particular importance for certain types of research (e.g., computer vision, speech and signal processing, scientific data reconstruction and reduction), where dedicated equipment and very high performance, but relatively inexpensive equipment, are all essential;
4. the extent to which *software technology* can and must support the move to large-scale parallel computing;
5. the developments in *communication technology* that will define the ways in which supercomputers can be made available to the U.S. research community;
6. the *efforts abroad* that might undercut the dominant U.S. role in high-performance computing;
7. the relationship of supercomputing to the broader field of *artificial intelligence;* and
8. the bottom-line technical policy question of how to get the U.S. research and industrial communities to *move forward in rapid and effective cooperation.*

II. CHIP TECHNOLOGY

The astonishing exponential growth of microchip performance is both the main driving force for the computer industry and the technical basis for expected parallel supercomputer developments. Though many circuit families of differing speeds and cooling requirements play a role in this growth, the following remarks will concentrate on just one family, the so-called field-effect transistors (FETs), which generate 40 times less heat than the faster bipolar chips, a very significant advantage as the number of circuits per chip increases steadily. State-of-the-art FET chips as of 1984 are represented by the 256K memory chip (512,000 devices), just now coming into mass production, and the roughly 100,000-device Motorola 68000 microprocessor. The 5 to 1 circuit density advantage of memory chips results from their very regular internal geometric structure and highly optimized design. Since 1960 these circuit densities have doubled roughly every 18 months, and this growth is expected to continue, with memory and microprocessor densities remaining in a similar proportion. By 1988 this should make 2-million-bit/chip memories available and could make million-transistor single-chip processors possible. As circuit sizes shrink, speeds will rise, so by 1988 the cycle-time of single-chip FET processors should have fallen from the 80 nanosec now typical to 30 nanosec for chips operated at room temperature and 15 nanosec for chips operated cryogenically. A typical application of this technology might be a single-chip, 64-bit scientific processor with full floating-point capability, on-chip memory management, internal overlap of instruction execution, 32 Kbytes of on-chip cache memory, and the capability of performing 20-40 MIPS. Processors tailored for somewhat faster execution of higher-level languages such as LISP and special-purpose computing devices like those described in Section V below (e.g., specialized vision-oriented or switching chips) all can be supported by this same technology.

The use of new materials such as gallium arsenide and other techniques for attaining higher electron mobility should allow perhaps 10 times higher performance, but they require significantly more difficult process and materials technology. The Japanese HEMT and the U.S. VHSIC are examples of such projects. Josephson circuits, which operate at cryogenic temperatures, and 3-D chips, which try to build up circuits in the third dimension rather than using the essentially planar structures now common, are not likely to be useful technologies in the near future.

Wafer-scale integration, including full and partial wafer techniques, have been considered for reducing the costs associated with the currently used wafer-dicing and reassembly technology. This approach faces severe chip yield problems, and one must reckon with the unfavorable signal-propagation characteristics of long on-chip conductors. For this reason, a packaging technique like that used in the IBM Thermal Conduction Modules, i.e., bonding individual chips to a multilayer ceramic substrate, may dominate for high-performance devices. Assuming that 200–400 chips are bonded to a single 6″ × 6″ ceramic board of this type, two such boards might by 1990 produce a 64-processor, 1,200 MIPS peak-rate parallel computer at a manufacturing cost between $50 thousand and $100 thousand.

Attaining the chip performances described in the preceding paragraphs will require attention to a challenging mass of supporting technologies, including special materials development, x-ray, E-beam, and ion lithography, as well as radiation damage removal methods. This work is moving rapidly in the United States and at least as rapidly in Japan, where experimental and fabrication facilities sometimes bettering the best U.S. facilities have been put in place. Japanese work in this critical area is spurred by a particularly clear national determination to succeed, resulting in a willingness to try approaches that U.S. workers tend to avoid because of initially unclear market perspectives. In this determination, combined with the dedicated large teams and engineering excellence brought to bear, lies the technological roots of the Japanese computer industry's rapid advance.

III. DISK TECHNOLOGIES

For supercomputers to be used effectively, they must be supplied as balanced systems, with input-output (I/O) bandwidth, external storage, displays, and other capabilities all appropriately matched to the expected two- or three-orders of magnitude increase in computation rates. Concern has been expressed about the ability of storage disks, a key element of this "peripheral" technology, to keep up with such rapid speed increases. The workhorse disk for current supercomputer systems is still the CDC Model 819 disk, a rotating storage device dating from 1970, with a capacity of 600 million bytes and a 5-Mbyte-per-second transfer rate. Two newer disks just now becoming available provide roughly double this capacity and storage rate, an exceedingly modest increase compared to far more rapid computation-rate speedups. Other more ambitious magnetic disk alternatives include the possible use of synchronized spindles or multiple heads in parallel. Although no such higher performance products have been announced (probably because of less-than-satisfactory past market experience with very-high-performance disks), products of this type may appear as derivatives of disk and head technology developed for larger, more lucrative markets.

Against this background, the development of a new optical disk storing 10^{11} bytes, now in progress under NASA and USAF (RADC) sponsorship, appears extremely significant. This disk, a first version of which should be available in 1987, promises to support a 10^8 byte/sec transfer, much better matched to anticipated supercomputer requirements than anything currently available. Access time, i.e., the time needed to start the inflow of a randomly selected data record, is anticipated to be 50–100 milliseconds. The development of this potentially important device ought to be pushed rapidly to prevent an initially limited market from delaying its appearance.

IV. CURRENT AND NEW COMPUTER ARCHITECTURES

During the next 5 years, new types of machines based on large-scale parallelism, having the ability to perform thousands or even tens of thousands of arithmetic operations simultaneously, will begin to appear alongside high-speed sequential and vector computers that dominate today's high-end computer market. At a rate determined by their performance, their cost, and their suitability for various applications, these new machines may progressively supplant present designs in some market sectors. Though our panel did not make any precise predictions concerning the rate at which this development will unfold over the next decade, we can describe the main architectural considerations relevant to this process.

High-speed single-thread machines keep programming relatively straightforward by behaving externally as if they execute instructions serially, even though internally a small number of operations (5–20) are actually being handled simultaneously. Machines of this design are the staple of present medium-high performance (4–20 MIPS) computing. It is technically possible to raise the performance of these machines to the 80–150 MIPS range with current technologies and to 250 MIPS or somewhat more using more advanced technology such as that represented by GaAs. Performance improvements beyond this will depend on the parallel use of multiple high-speed serial processors. Very-high-speed serial machines and multiprocessor configurations involving small numbers of such computers are bound to retain major, perhaps dominant, economic importance for the next 5–10 years, especially for commercial applications. In the supercomputer environment, fast (4–10 MIPS) serial computers will be the single-user "workstations" that supply researchers with their software development tools and much of their graphic and data display environments.

Vector supercomputers, like the present U.S. Cyber 205, CRAY XMP, and the Japanese Hitachi S-810/20 and Fujitsu VP-200, can be regarded as simple parallel machines specialized for the handling of computations characterized by particularly regular patterns of data motion. Such regularities typify certain important numerical scientific codes, and hence fast vector machines are generally designed as "scientific number crunchers" equipped with the highest-speed floating-point operation units. The current peak performance of these machines is in the hundreds of MIPS. Single-processor machines of this type could achieve operation rates of 250–500 MIPS with the best present technology and perhaps twice this using GaAs technology. Multiprocessor configurations of vector processors, a type of design under active consideration by several supercomputer manufacturers, will probably attain peak performances of roughly 10,000 MIPS* using 16–32 processors within 5 years, rising to 25,000 MIPS within the coming decade. At such performance levels, they can be expected to retain a major role, and perhaps a dominant role, in bread-and-butter scientific number crunching through much of the decade. Note, however, that computers of this design can exploit parallelism only in innermost program loops; and for programs that branch or address memory in irregular patterns, the actual efficiency typically decreases to 15–20 percent of rated peak performance, e.g., for such problems a 4,000 MIPS actual computation rate is all that can be expected from a machine nominally rated at 25,000 MIPS.

New parallel supercomputer designs, many of which are still highly experimental, are at once the focus of much current research and the one area (of design, as distinct from technology) in which sudden surprising advances might occur. Many of these designs promise large increases in computational power without significant redesign of their components, i.e., they are "scalable." With such architectures it may be possible to increase machine speeds by factors of 10 or even 100 by increasing the amount of parallelism used, without major redesign but at a proportional increase in cost. Once suitable software has been developed for machines of this type, a major research undertaking, the homogeneity of their structure should make it possible to move applications to much larger-scale parallel machines with minimal software change. Moreover, scalable parallel machines have major potential

*The performance of high-speed scientific computers is often rated not in terms of the total rate at which instructions of all types are performed but only counting that part of the total activity devoted to the execution of "floating-point" arithmetic operations: MFLOPS (millions of floating-point instructions per second) rather than MIPS (millions of instructions per second). The ratio between these two measures is typically taken to be 1/3, but since this ratio is highly application dependent, we have chosen to sidestep this distinction and to state all crude performance figures in MIPS.

advantages in a VLSI environment because they are built from large numbers of identical parts that can be mass-produced efficiently and because the design cost can be amortized over multiple configurations differing only in size. For all these reasons, we review this class of computers at length and attempt to clarify the technical options available for parallel supercomputer development by listing various principal design alternatives characterizing the design of such machines.

Most of the more than 70 parallel machines that have been proposed by university and industrial groups here and abroad occupy easily characterized positions in the "design space" defined by the following alternatives:

1. *SIMD versus MIMD.* An SIMD (single-instruction multiple-data) machine consists of multiple processors (or of a few very fast pipelined processing elements), all of which are fed synchronously by a single instruction stream but which operate on independent streams of data. Vector machines like the CRAY I or Hitachi 810, and synchronous array machines like the ILLIAC IV and NASA/Goodyear MPP, typify this class. An MIMD (multiple-instruction multiple-data) computer consists of multiple processors (or fast time-sliced processing elements) driven by independent instruction streams and capable of branching independently but also able to pass data and synchronization signals between processors, perhaps via a shared memory. The Denelcor HEP, Cal Tech Homogeneous Hypercube, and numerous other machines belong to this class. SIMD machines are generally more efficient for applications characterized by very regular patterns of processing, in part because their "lockstep" mode of operation shares a single unit of control hardware among many data-processing elements and eliminates much of the message-passing overhead needed when all the processors of an MIMD machine must proceed through a computation in close synchrony. On the other hand, it is hard to make SIMD computers deal effectively with highly "branched" code, since the process of branching in such computers is normally imitated by temporarily disabling some of the processors in the parallel array, and thus several successive branches tend to reduce effective computational power drastically. MIMD processing arrays deal more robustly with complex branched code, since they are capable of executing not only innermost loops, but even complex branch-filled loops, in parallel.
2. *Visible versus "under the covers" physical communication pattern.*

In all parallel machines some portion of the computation involves communication among the individual processing elements. From the physical point of view such communication is allowed in only a few specific patterns defined by the hardware. For example, the processors may be arranged in an obvious two-dimensional grid with each processor connected to its north, south, east, and west neighbor. A single operation on a machine wired in this pattern can only send a number from each processor to one of these neighbors. Proposed connection patterns for parallel machines include rings, n-dimensional cubes, binary trees, and various less obvious but much more powerful communication networks developed by communications researchers.

Once having chosen a communication net appropriate for the computational structure being developed, a parallel machine designer can choose either to make the structure of this net visible to the machine's user or to hide these details behind a set of programmed conventions that makes it appear that any processor can communicate directly with any other.

Both alternatives have arguable advantages. On the one hand, if a machine's actual communication pattern is well matched to an application and is made available to the programmer, it can be exploited to gain speed. Examples are the use of two-dimensional grid patterns for image processing and the use of various powerful communication patterns for optimal implementation of the Fast Fourier Transform. On the other hand, machines that relieve their programmer from direct concern for the pattern in which messages must flow between processors are easier to program for a wide range of problems, particularly problems characterized by highly variable patterns of communication. Another potential advantage of machines in this latter class is that the communication paths that they use, being hidden, can be varied dynamically, e.g., to bypass faulty components.

3. *Coarse-grained versus fine-grained designs.* In any parallel computer with multiple processing elements there is a cost trade-off between the number and the size of the processors. The approach that stays closest to established single-stream processor technology is to use as few as possible of the largest available processors. The opposite approach is to achieve as much parallelism as possible by using a large number of small elementary processors, which can be single-chip microprocessors chosen to exploit the cost-effectiveness of a mass-produced item or, in the limit, can be a

single-bit processor used in very large numbers. The decision to use a relatively small number of very-high-speed processors is justifiable by the argument that even largely parallelizable code will have significant serial sections, so high performance in serial code will be important for sustained overall high performance. Its polar opposite, the use of many single-bit processors, can be justified by a desire to push parallel execution to its limit, and by the surprising level of arithmetic power per square millimeter of silicon that one-bit processors attain. Perhaps the most important issue here is one of programming style. Since ordinary serial machines are coarse-grained, the technology for programming coarse-grained machines is understood best. Thus, it is plausible to expect a FORTRAN compiler to optimize code that keeps, say, 16 processing units busy, but not 16,000. On the other hand, if an algorithm is written with parallel processing in mind from the start, it may divide naturally into numerous processes, e.g., the processing of a 1,000 by 1,000 image in which relatively simple, low-precision algorithms must be applied to individual pixels or small pixel groups may fit most naturally on a million processor machine.

4. *Specialization of architecture to an anticipated application or programming style.* Scientific number crunchers are optimized for arithmetic operations on floating-point numbers, while processors intended for more varied symbolic applications must be optimized to handle irregular memory references and branching patterns, fast subroutine transitions, and fast context switching. They must provide run-time type checking and must support efficient execution monitoring during program debugging. Thus in designing a "scientific" processor one will emphasize certain types of hardware that are of little use in a "symbolic" machine, and vice versa. Various designs for large-scale "dataflow" computers, which are specialized in a different direction, illustrate this same general point. Dataflow programming is organized around a set of fully parallel concepts. A program is regarded as an assemblage of "operators," all poised to fire as soon as they receive all their logical inputs. As soon as it fires, each such operator transmits its output to further operators, for which it becomes one of several possible inputs.

The innermost interpretation cycle for such a language consists of various operand-routing, instruction-dispatching, and instruction-execution activities, many of which can be overlapped.

As compared to a general-purpose parallel assemblage, a cleverly designed layout of dataflow operation nodes in memory and effective instruction dispatch hardware can probably double or triple the efficiency with which a dataflow language executes, while adding only minimally to the total hardware cost of a parallel system. These remarks typify the speed gains achievable by the architectural specialization of parallel machines.

V. SPECIAL-PURPOSE COMPUTING DEVICES

Extremely high computation rates can be attained efficiently by tailoring electronic hardware to algorithms of special importance (e.g., the Fast Fourier Transform) or to the requirements of a particular computer-intensive problem. Devices of this sort are of particular importance to diverse cutting-edge research efforts and to applications where dedicated equipment is essential but dedicated use of multimillion-dollar general-purpose computers such as a large vector supercomputer would be prohibitively expensive.

These special-purpose computing devices point to a class of computing devices highly optimized to particular applications, which might be made possible by hypothetical future advances in compiling technology that allow automatic generation of whatever computing device is best suited for any specified application. Though this is far beyond present capabilities, manually designed special-purpose computing systems are of growing significance to defense uses, such as signal processing and image analysis, to high-energy-physics data acquisition and reduction, and to other areas.

It is steadily becoming easier to realize such systems, either by combining relatively standardized high-performance modules into specially wired configurations or by custom gate-array or VLSI designs. Special systems produced in this way tend to reflect end-user performance requirements closely, and hence they uncover important subproblems of the general supercomputer design problem quickly. Moreover, because of the emphasis on cost-effectiveness typically characterizing their design, they encourage industry to improve the effectiveness of more general computing engines supplied commercially. Also, though machines of this sort sometimes involve approaches that address only one specific problem, others affect larger classes of problems in an extremely cost-effective way, making them desirable attachments for general-purpose supercomputer systems. Encouraging a steady flow of such designs, which sometimes advance the effective frontier of computing in the application areas they address, can strengthen the U.S. competitive position significantly.

The sum of these special design efforts has become an important component of computer research that needs to be explicitly recognized and systematically cultivated. For this, an improved infrastructure for rapid and effective development of special-purpose systems is a prime necessity. In particular, projects motivated by special problems need a basic kit of standard modules out of which optimized architectures can be built. This should include low-cost single-board computers, high-density multiport memories, and high-speed switching modules that allow large communications networks tailored to specific applications to be fabricated rapidly. Some of these items are beginning to appear commercially, but further encouragement to industry and systematic attention to standardization issues are appropriate.

To increase the speed with which prototype special-purpose computational designs can be developed for significant experiments, system-level tools of the same quality as the best current VLSI design tools need to be developed. A focused program to develop such tools and make them available to universities, laboratories, and private-sector organizations appears to be an effective way to strengthen the work of special-system designers. Such tools would also facilitate the transfer of successful prototypes into industrially supported products, which is as important for special high-performance computational designs as for general-purpose supercomputer development.

VI. THE ROLE OF COMMUNICATION TECHNOLOGY

Supercomputers communicate three ways: internally, locally (in the same building), and outside (to machines elsewhere in the world). Existing technology for internal connections uses highly parallel buses and is limited primarily by wire length. Existing nearby machine telecommunications technology runs at a maximum of about 200 Mbit/-sec, which is faster than most individual computers can accept information. Current outside connections are limited by available nationwide channels to 3 Mbit/sec, and 50 Kbit/sec is more typical.

In the next 5–10 years it is unlikely that the basic internal communication technology will change much. However, it is possible that within 20 years optical connections or wafer-scale integrated circuit technology will have significantly changed the way computers communicate internally.

Local interconnections will be affected by optical technology much sooner. It is possible that within the next 5–10 years commercial optical local networks running at a gigabit/sec rate will become available. This rate will not be accessible to any individual machine but will permit many supercomputers to share a relatively inexpensive network and exchange data at their internal bus rates.

NBS, DARPA, and cooperating international standards organizations have made large-scale networks possible by specifying protocols and standards that promote interoperability of heterogeneous equipment from different manufacturers. However, nationwide network speeds are currently inadequate for the amount of data consumed by supercomputers. Over the next 5 years the NSF Science Net project should be in place with high-speed satellite connections linking the major research universities. This will be a significant help to providing access to supercomputers.

Finally, we note that communication security on these nets may become an important issue, and this area deserves careful and early attention.

VII. ALGORITHMIC AND SOFTWARE ISSUES

Effective parallel processing will require a fundamental understanding of computational models, algorithms, programming languages, optimization techniques, processor-memory-I/O relationships, and interfaces to existing computing systems and end users. Though research on parallel algorithms is expanding, the lack of large-scale parallel machines for experimentation orients current work toward the study of individual algorithms rather than on practical applications, systems questions, or large software structures. Moreover, systematic work on large algorithm libraries, such as would quickly develop around major parallel computing installations, is hardly possible in the present atmosphere of largely theoretical studies. Also, although operating systems for parallel machines raise many subtle problems, it is difficult to address these except by working with real parallel systems.

Parallel programming is much more complicated than sequential programming because it deals with simultaneously occurring events involving potentially complex (time-dependent) interactions, so that elusive and nonrepeatable bugs can occur. For this reason, the usability of large parallel machines will be bound up with the development of programming languages that facilitate the expression of parallelism and the development of sophisticated optimizing compilers that improve the execution efficiency of such languages and find implicit parallelism where it is not explicitly expressed.

Hence, although initial experience (e.g., work on the HEP at the Los Alamos Laboratory) indicates that modest extensions to conventional languages will allow important applications to run with high efficiency on certain types of parallel computers, many new programming language concepts need to be investigated as ways for easing the further exploitation of parallelism. These include languages with vector and set operations (e.g., APL), dataflow languages, applicative languages, languages such as parallel LISP and PROLOG that emphasize recursive branching and backtracking processes, and languages in which communicating sequential processes play a central role. It is important both to encourage theoretical study and development of all these classes of languages and also to ensure that they become available for extensive experimental use as soon as substantial parallel computers suitable for them are constructed.

The present state-of-the-art software to support high-speed parallel computation is also represented by vectorizing compilers and more general parallelizing compilers capable of finding certain opportunities for parallelism automatically by analyzing ostensibly serial code. Both Cray Research and Fujitsu Ltd. have produced such compilers for their vector supercomputers, as have CDC, and U.S. research projects at the University of Illinois and Rice University.

While very useful, these systems are not yet able to automatically translate every standard sequential FORTRAN program into a good parallel program. Indeed, this will remain beyond the capabilities of automatic systems in the foreseeable future because the best parallelization usually requires substantial comprehension of the underlying problem domain. However, these systems do provide invaluable aids to the programmer because they permit him to write parallel programs cleanly in a high-level language without cluttering the language with special-purpose constructs.

Present automatic code vectorization techniques need to be evolved to handle programming for parallel multiprocessor supercomputers and to apply to whatever new parallel languages appear most important. Many of today's vectorization techniques are directly applicable; others will need to be adapted substantially or new methods invented. Program development environments tailored to the special requirements of new parallel systems and languages also need to be developed, especially to aid in the debugging of parallel programs. Graphics systems making it easy to deal with the complex outputs that parallel supercomputers can generate will also be needed.

Complex software developments cannot proceed far in a purely "paper mode." For them to flourish, real and relatively large-scale experiments with functioning hardware are essential. Thus, to accelerate the very challenging software research outlined in the preceding paragraphs, it is essential to deploy or provide remote access to experimental parallel machines as soon as they become available. Unconfining access is essential to the major systems software research and development efforts that will be required to ensure effective use of these new machines.

VIII. EFFORTS ABROAD

U.S. supercomputer developments must be appreciated in the context of efforts abroad, primarily those of the powerful and fast-moving Japanese computer industry, and also relative to an accelerating European effort. Our panel's conclusions concerning the Japanese program are as follows: The largest Japanese vector supercomputers (Hitachi S-810/820 and Fujitsu VP-200) have attained parity with, but not surpassed, corresponding U.S. machines such as the CDC Cyber 205 and the CRAY XMP. Japan's stated national commitment to advance in this area, the great engineering strengths of Japanese industry, and the leading position that Japan has won in certain lines of high-speed semiconductor component production will make for a very closely run race in this area. Nonetheless, the United States can at least expect to maintain parity in vector supercomputer production, and moreover in various rapidly expanding U.S. supercomputer use-oriented educational and research programs, e.g., the newly announced NSF Supercomputer Access Program, will help preserve U.S. leadership in vector supercomputer applications development.

Competition in regard to highly parallel supercomputers of newer experimental designs favors the United States even more strongly. The United States can expect to retain world leadership in this area, provided that agency support focuses on the most robust new designs and that proper bridges are built between university-centered innovation and the organizational and manufacturing capabilities of the U.S. computer industry. Without such links, it is feared that competition in this area will find various relatively small U.S. firms, including some recent start-ups, forced to compete with the extremely impressive technological capabilities of such major Japanese firms as Fujitsu, Hitachi, and NEC. It is noted in this connection that some Japanese firms have evinced interest in exchanging technology and development support for the results of U.S. conceptual research.

Software construction is also an area in which Japanese strengths, though not yet widely recognized in the United States, are already impressive, as shown, for example, by the first-rate vectorizing compilers available with the new Hitachi and Fujitsu vector supercomputers. Though the technology involved derives almost wholly from U.S. work, a recent study of Japanese software engineering practice, done by the JTECH Panel on Computer Science under the sponsorship of the U.S. Department of Commerce, concludes that in the production of standard software systems Japan may already be ahead of the United States and may be getting further ahead. For example, the average software production per employee in the 2,000-employee Toshiba software factory (which produces complex distributed systems process control software for power plants, steel mills, flight guidance, and so forth) is reported at 2,000 lines of high-level code per month, as compared to a typical U.S. figure of 300 lines per month. Error rates of 0.3 errors per 1,000 lines are achieved, as compared to U.S. rates 10 times as high, which may provide Japanese vendors the competitive advantage of marketing software with a 10-year warranty. Much higher rates of software module re-use than characterize U.S. practice are attained. The cited JTECH Panel report ascribes these significant successes to the systematic application of disciplined coding techniques and to source code control methods, quite familiar in the United States but much less systematically applied, and also to the sense of responsibility for a product that pervades Japanese society and to the factory discipline that Japanese management is able to apply to software production. This Japanese prowess in an area far more central to the computer industry than supercomputer production can potentially erode U.S. systems sales significantly and calls for further prompt analysis and appropriate response.

In spite of the intense international attention and publicity that have been focused on Japan's "Fifth Generation" computer effort, it is not clear to our panel that the ambitious goals set for it are matched by any entirely persuasive technical idea of how they can be reached. Nevertheless, it must be understood that even if not fully successful the kind of ambitious long-range research represented by the Fifth Generation project will generate significant new engineering, architectural, and software strengths for Japan. It must also be understood that Japanese universities and firms have mounted many parallel-computer explorations smaller and less publicized than the Fifth Generation effort and that cumulatively these will position Japan to move forward rapidly with new parallel machine designs. It is quite possible that these new Japanese designs will emerge immediately after the current wave of U.S. effort has determined what lines of development are most promising.

We note also that it is clearly in the U.S. interest to build up a much larger body of computer specialists familiar with Japanese work and able to read Japanese technical literature. As long as tens of thousands of Japanese researchers follow the U.S. literature diligently while only a handful of U.S. computer scientists can deal with Japanese, the main flow of technical information will clearly run from the United States to Japan.

The European "Esprit" program does not appear to confront the United States with competition as severe as that seen in Japan. A review of the long list of projects to be supported under this program suggests that it will serve to strengthen European work in a broad range of technological areas, but concentrated work apt to result in supercomputer developments that will create major competition for the United States does not appear likely. Nevertheless, this European work needs to be followed closely.

The British response to the Fifth Generation challenge is the "Alvey" program, which is similar in its goals to the U.S. MCC research consortium now established in Austin, Tex. The four technologies making up Alvey are VLSI, software engineering, AI, and man-machine interfaces. Extensive university-industry cooperation is planned, and U.S. companies are welcome to participate as long as the research is done in Britain. Although the Alvey focus on supercomputer issues is not yet clear, it is important that this potentially significant national effort be followed closely.

IX. THE ARTIFICIAL INTELLIGENCE ISSUE

Construction of computerized artificial intelligences can justly be regarded as one of the most central long-term goals of computer science (and even of science generally), and it deserves tenaciously sustained support. The supercomputer developments now in progress will undoubtedly advance artificial intelligence research by putting new technical means at its disposal. An initially modest but growing degree of specialization in computer architectures can be expected to develop in response to the requirements of this field. Though the algorithmic structure of artificial intelligence computation is much less mature and stable than that of other areas of computation (e.g., numerical scientific computation or computer algebra), it is clear that work in artificial intelligence can be expected to emphasize random patterns of memory reference, various exact and approximate matching processes, computations with short quantities rather than multidigit floating-point quantities, and languages such as LISP whose execution can be accelerated (perhaps tripled) by suitable hardware. Moreover, the first stages of data input for subfields of artificial intelligence such as image and speech processing will require various special-purpose computational devices of the kind discussed in Section V, while exploration of various programming styles currently of interest to workers in artificial intelligence, e.g., fast interpretation of large systems of production rules, high-speed formula unification in logic, or neutral-net simulation, may also justify specialized hardware. The scientific computation and artificial intelligence communities favor distinct programming languages, emphasize different aspects of system performance, and prefer different operating systems, making it difficult for them to be served effectively by a single operational environment. Nevertheless, it remains difficult to derive any specific supercomputer designs from the requirements of artificial intelligence research, and in this sense the division of supercomputers into two distinct genuses of "scientific machines" and "artificial intelligence machines" is artificial. Perhaps it is best to say that artificial intelligence, as a young field in which approaches are still evolving, requires computers that can deal with rapidly shifting patterns of computation and communication, with great flexibility and at very high speed.

X. HOW TO MOVE FORWARD

Several major federal initiatives aimed at strengthening U.S. supercomputer-related capabilities have been launched and, if pursued along the strongest technical lines and with due regard for the importance of early and effective industrial involvement, can maintain U.S. computer preeminence. Principal among these are the DARPA Strategic Computing Program, the expanded supercomputer research programs of the

Department of Energy, and NSF's Supercomputer Access Program. The steady support that NSF has provided for many, indeed most, of the university groups whose work has defined the principal supercomputer architectural alternatives now available has been of fundamental importance and is planned to continue and even to increase modestly.

Since DARPA's architecture program comprises the largest single federal project addressed to supercomputing, its activities will be broadest and most crucial. Work in many areas, including underlying technologies, improved design automation tools, fabrication means for both VLSI and broad-level special systems construction, conceptual exploration of new special general-purpose parallel computer systems, various small-scale prototyping efforts, and a few critical large-scale prototyping efforts, will all be sponsored.

Our panel is supportive of DARPA's intent to organize the selection of the culminating major efforts of this multiyear program via an orderly series of phases, beginning with conceptual studies aimed at the discovery of significant new computer architectures, progressing to the exploration of those architectures by simulation and experimental software development, then to small-scale hardware prototyping where justified, and finally, in the most promising cases, to the construction of large high-performance prototypes. We note that for these large culminating prototypes, research success, and even more the ability to move successful research results into broad U.S. use before foreign competition follows suit, will clearly be dependent on serious industrial involvement. The proper use of academic resources is in research and small-prototype development; the realization of large systems requires the skills of proven engineering organizations with established track records. For these reasons, we recommend that both DARPA and DOE encourage or require substantial industrial involvement (by either established or start-up companies) in the development of each experimental computer system chosen for large-scale prototyping. We understand that it is DARPA's intent to encourage this type of industrial involvement and we support this decision. Ideally, this should involve a substantial commitment of capital and crucial technical skills by the company (or companies) cooperating in such a development, to ensure that careful and independent private-sector technical judgment concerning project soundness is brought to bear in each such case. Moreover, to strengthen university-industry links and overcome what might otherwise be substantial familiarization delays in moving promising designs into large-scale prototyping, it is recommended that both DARPA and DOE encourage industrial involvement even in the initial smaller-scale stages of parallel computer prototyping. Early industrial involvement in special-purpose computer developments is also desirable.

Though considerably smaller than DARPA's effort, the DOE supercomputer program has maintained a significant focus in areas complementary to, and synergistic with, the DARPA effort. Historically, DOE has been the leading agency for the applications of large scientific computing and has maintained particularly close relationships with the supercomputing community. Since optimization of computer architectures for floating-point computation and for the regular patterns of data manipulation characterizing many numerical applications may demand significantly different designs than are optimal for nonnumerical applications, it is important to maintain a strong program of architectural research and development with a clear scientific-computation focus. DOE's strong tradition and high level of expertise in scientific supercomputing make it appropriate for DOE to play a significant funding role for research on advanced computer architectures aimed at high-performance numerical computing, and acceleration of its program in this area is desirable. DOE is also in the process of significantly expanding its supercomputer access program to include scientists in High Energy and Nuclear Physics, Basic Energy Sciences, and Biological and Environmental Research. This expansion will contribute substantially to solving the national supercomputer access problem. In addition, DOE has acquired several commercial parallel computers and is making them available to many researchers. These DOE programs have played and will continue to play a very important role in maintaining overall supercomputer application capabilities.

As currently defined, NSF efforts will concentrate on making supercomputer capabilities broadly available to the U.S. scientific community, including but not restricted to computer scientists. Though initially this program will concentrate on providing access to existing vector supercomputers, our panel sees the NSF Access Program as having broad continuing implications and recommends that NSF plan to make all major experimental parallel computers nationally accessible as soon as feasible. Efforts of this sort, which serve to build supercomputer-use capabilities, strengthen the U.S. position by exploiting one of our most characteristic advantages, the great breadth and diversity of the U.S. computational community. However, the resources of this community will be heavily taxed by the deep algorithmic and software problems associated with the new classes of machines that will be developed over the next decade. Training of an expanded population of workers in this field must therefore be seen as an essential component of a balanced national effort.

The NSF Supercomputer Access Program is well calculated to serve the requirements of the large U.S. scientific community for whom computers are a tool for other research goals. However, in view of the great importance that NSF basic computer science support has had in the development of new architectural alternatives and the related computer science until now and in view of the excellent technical judgment with regard to hardware and software that NSF can bring to bear, the NSF programs should also be structured to support major software and applications development for the machines to which access is provided. Moreover, to avoid dissipating the attention of research groups among overly many new machines, all the agencies sponsoring supercomputer research and development should plan to support groups wishing to develop algorithms and software for experimental hardware systems produced commercially or by other universities. Such participating research groups should be guaranteed adequately "privileged" remote access, e.g., the right to take full control of a system for scheduled blocks of time to allow exploration of operating system changes in the light of new research ideas. In some cases, it will also be appropriate to disseminate smaller-scale versions of major experimental systems to several locations.

The need for quantitative measurements of progress and quantitative characterization of the advantages of alternative architectural approaches underlies all the areas of research we have discussed. It is therefore appropriate to encourage joint industry-university-government development of tools and techniques to measure progress in software, architecture, and algorithms for supercomputing systems.

XI. CONCLUSION

The U.S. national interest in the rapidly moving area of computer architecture should be as follows:

1. To ensure that all designs likely to succeed in any major way are explored actively. We must guard against technological surprise.
2. To get successful, or potentially successful, designs into the hands of large user and software development groups quickly, so as to accelerate the development of the necessary software. We must maintain system and software leadership.

3. To accelerate the transition of successful designs into commercial production, preferably by involving strong industrial groups in their development from the start.

It is quite important for U.S. activities aimed at maintaining awareness of Japanese technical developments through suitable translations be accelerated, possibly through programs undertaken by NSF or DARPA.

Finally, since the U.S. faces rapidly mounting competition in this area from both Japan and Europe, it is important that vigorously conceived programs, including those now planned by the major agencies involved, move forward without delay.

RESEARCH BRIEFING PANEL ON INFORMATION TECHNOLOGY IN PRECOLLEGE EDUCATION

John Seely Brown ***(Co-chairman),*** *Xerox Palo Alto Research Center, Palo Alto, Calif.*

James Greeno ***(Co-chairman),*** *Graduate School of Education, University of California at Berkeley, Berkeley, Calif.*

Henry Becker, *Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Md.*

Richard Burton, *Xerox Palo Alto Research Center, Palo Alto, Calif.*

Allan Collins, Bolt, Beranek, and Newman Laboratories, *Cambridge, Mass.*

Andrea diSessa, *Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Mass.*

*Harry Lasker, *Interactive Training Systems, Cambridge, Mass.*

Mark Lepper, *Department of Psychology, Stanford University, Stanford, Calif.*

*Dr. Lasker was unable to attend the meeting of the briefing panel, but met separately with the co-chairmen. ***Additional advice was obtained from: Alan Lesgold, Lillian McDermott, Arthur Melmed, Andrew Molnar, Jack Palmer, Roy Pea, Senta Raizen, Paul Resta, Edward Silver, Charles Smith, John Tukey, and Fred Weingarten.***

Milbrey McLaughlin, *Graduate School of Education, Stanford University, Stanford, Calif.*

Robert Parry, *Department of Chemistry, University of Utah, Salt Lake City, Utah*

Henry Pollak, *Bell Communications Research, Inc., Summit, N.J.*

Frederick Reif, *Department of Physics and School of Education, University of California at Berkeley, Berkeley, Calif.*

Staff

Norman Metzger, ***Staff Officer***

Susan Newman, ***Rapporteur,*** *The Learning Lab, WNET-TV Education Division*

Gerry Kasarda, ***Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON INFORMATION TECHNOLOGY IN PRECOLLEGE EDUCATION

SUMMARY

The development of more effective research methodologies and tools for the study of human learning, as well as recent advances in hardware capabilities, is opening up the possibility of using powerful new information technologies to aid individual learners and to help solve pervasive educational problems. Parallel work in artificial intelligence and the cognitive sciences has set the stage for qualitatively new applications of technology to education. What is required to move forward is increased support for basic interdisciplinary research, focused by the development of advanced learning systems employing the methodologies and equipment of artificial intelligence.

While success in this arena will not solve all educational problems, it will provide a new scientific basis for instructional and systems design, teacher training, and curricular restructuring. It also will create valuable new resources in the form of model electronic learning environments, while attracting a new cadre of professionals to the fields of education and educational research.

Thus, the panel recommends two federal initiatives to improve

precollege education through the use of technology: (1) Support increased basic cognitive research on learning undertaken in conjunction with the development of advanced electronic learning environments employing the methodologies and equipment of artificial intelligence. (2) Continue federal support for individual scientists and applied research and development centers to analyze specific educational problems, develop and evaluate educational software, and find effective ways to integrate educational technologies into classrooms.

INTRODUCTION

Numerous recent reports have cited problems in the schools, declines in student performance on standardized tests, and the weakness of American mathematics and science education relative to other industrialized countries. Diverse sectors of society have begun to call for increased educational effectiveness and, in particular, for exploring the potential of information technologies for improving education.

However, earlier promises of educational improvements via technological interventions, such as educational television and computer-assisted instruction, have not always been fulfilled. Current investments in microcomputers are seen as limited in their effectiveness. Therefore, one charge of the present panel has been to assess what is different now—technologically, scientifically, and sociologically—that warrants increased federal attention to the research and development of instructional systems for the improvement of precollege education and what is required to advance the potential for learning represented by information technologies.

WHAT IS DIFFERENT?

While past projects to reform education based on learning research and technological advances have not always fulfilled the hopes of their designers, they have resulted in limited advances. Moreover, they have provided opportunities to learn how to achieve needed change more effectively. Recent studies of past reform efforts have identified specific difficulties surrounding school change and recommended how such difficulties can be surmounted. In addition, a confluence of several other factors is creating a unique opportunity to enhance the value of information technologies in precollege education.

NEED FOR EDUCATIONAL CHANGE

While observers are alarmed by evidence of declining "basic" mathematics and literacy skills among American students, there is simultaneously a growing emphasis on teaching "higher-order thinking skills"—reasoning, critical thinking, and learning about learning. These objectives require increased pedagogical emphasis on integrating basic skills with conceptual understanding, developing critical thinking and reading skills, enhancing the ability to see connections among objects and processes, and developing the capacity for self-directed learning.

RAPID DISSEMINATION OF COMPUTERS

The growing ubiquity of computers, while having at present relatively little direct educational impact, is nevertheless familiarizing a generation of teachers, parents, students, schools of education, school boards, and others with computers as tools and media. This sort of "computer literacy" sets the stage for the acceptance of more advanced educational applications such as those outlined in this report.

In addition, dramatic increases will occur in the computational power available to consumers at affordable prices within the next 5 to 10 years. These new levels of machine power make possible the delivery of the kinds of integrated and knowledge-based systems that are foreshadowed by present research and development efforts but that require concentrated basic research and experimentation for their full realization.

EFFECT OF TECHNOLOGICAL DEVELOPMENT ON RESEARCH

Developments in the cognitive sciences and advances in computer technology have had a deep impact on educational research. Prior to 1960, learning research was concerned with primarily discovering the external conditions that influence readily observable behavior. While providing useful indications of the limits of human cognitive capacities and a scientific basis for the field of instructional design, behavioral psychology yielded little insight into the "black box": the underlying cognitive events that mediate behavior, including learning.

However, during the last 25 years, researchers have gained a new language and set of tools for expressing and validating learning theories that focus specifically on internal cognitive processes and their relation to the growth of knowledge and mental operations in humans. For example, symbolic computing languages used by the artificial intelligence community, such as LISP, have provided a new notational system that is both scientifically precise and sufficiently flexible to represent complex learning processes.

In addition, computer technology has brought important changes in educational research by acting as a finely tuned "cognitive lab" for testing hypotheses and manipulating the large number of variables affecting human cognition. Through the simulation of human learning and problem solving, researchers have been able to verify detailed hypotheses about the mechanisms, architecture, and language of the human mind, particularly as they affect highly skilled performance by experts or the mental work by which novices gain new skills.

Finally, owing to the rapid growth of the information technology industry, a new generation of college and graduate students is becoming well versed in the computer sciences. The result is a larger pool of scientifically trained individuals from which, given adequate incentives, it is now possible to draw talent capable of conducting the combined educational and technological research needed to address new educational objectives.

GOALS AND OPPORTUNITIES

In addition to traditional competencies, students increasingly should learn the thinking skills to manage information, formulate effective probing questions, test hypotheses, make judgments, express themselves logically and lucidly, and solve problems.

Unfortunately, our present educational resources are inadequate to meet these objectives. In most classrooms, there is a lack of integration between low-level skills and higher-level understanding of a subject, and between disciplines such as mathematics and science, or writing and other subjects. Cognitive research confirms that knowledge and skills learned without conceptual understanding or functional application to problems are either forgotten or remain inert when needed in situations that differ from those in which they were acquired.

This common dissociation of formal schooling from experiential learning prevents students from mobilizing for formal learning the power of their intuitive understanding, even though they use it effectively outside the classroom. Students receive feedback in the form of grades, but are given little opportunity to revise their school work or to reflect on or communicate with other students or the teacher about their own writing, problem-solving, or other learning efforts. As a result, many students fail to correct even minor misunderstandings that nevertheless prevent continued advancement, and to develop skills for assessing their own learning problems.

While many of the needed changes could be effected by human teachers, appropriately trained, paid, and teaching smaller and fewer classes, the costs of this approach appear to be prohibitive. The possibility exists, however, that information technologies can provide the means to deliver some aspects of instruction not only cost-effectively but also in a deeply individualized and integrated way.

Specifically, parallel research in cognitive science and artificial intelligence is within reach of developing intelligent diagnostic, coaching, and discovery learning systems for a variety of subjects. Such systems could have a major impact on education. For example, knowledge-based simulation systems can enable students to test their understanding of scientific and mathematical principles, posing original problems that reflect their own questions about the world. With more powerful computers and research, it will be possible to create sophisticated work stations that can be used over a student's school lifetime. Such systems might include computerized lab-simulators, data-analysis software, writing and idea-structuring tools, and problem-analysis and management software.

In short, the real promise for the future is the development of integrated systems capable of interacting intelligently and sensitively with students and of providing a responsive, structured environment in which various skills and concepts can be developed and practiced. To realize this potential, initial investments must be made to capitalize research equipment and to support basic artificial intelligence and cognitive research focused on education and the development of educational technology. Such investments will enable us to move beyond the usual ad hoc applications of technology to educational problems. Besides providing a better basis for curriculum design, organization of subjects, and teacher training, increased scientific research will enable us to tailor instruction for individuals to a degree not generally found outside the apprenticeship learning models prevalent in graduate schools or other research centers.

TECHNOLOGY IN PRECOLLEGE EDUCATION

Information technology applied to education can be divided into three levels: systems of hardware and software now commonly used in schools; systems currently in the development and commercial pipelines; and advanced experimental systems embodying cognitive science and principles of artificial intelligence.

LEVEL 1: COMMON USES OF TECHNOLOGY

Most of the computers presently available in schools are 8-bit microcomputers, with limited graphics, memory, and processing capabilities; however, some elementary and high schools are purchasing 16-bit and 32-bit "personal computers." While a growing range of applications is being developed, the majority of the schools that have microcomputers now use them to teach "computer literacy," BASIC programming, and for drill-and-practice or other forms of programmed instruction.

The best versions of the latter are based on behavioral analyses of the components of skill required to perform a given learning task, together with research on simple motivational factors and reinforcement schedules that influence persistence in a task. These types of programmed instruction are those most easily implemented on computers. In addition, these systems provide students with the advantages of sequenced practice, with the problems and text material adjusted to the level of student performance. Gains in student performance of routine skills, such as arithmetic calculation and elementary programming, have been accomplished through the use of these systems.

Assessment

While providing efficient means of acquiring rote skills, the educational benefits to be gained from these baseline resources are likely to be quite modest. The limited processing power and memory of microcomputers preclude their use for learning activities that conceptual understanding with mastery of procedural skills more effectively integrate; further, the analyses on which the programs are based do not address the cognitive mechanisms by which this sort of integration can be achieved. Computer literacy courses are rapidly outmoded as new hardware and software are developed. In addition, a new generation of more powerful and flexible programming languages and techniques is being created. As a result, instruction in BASIC may provide a poor model for students who might wish eventually to tackle complex programming problems.

LEVEL 2: RECENT INNOVATIONS

More imaginative educational uses of computers are being developed commercially or in cooperation with university-based researchers. These applications, which are being used in homes and experimentally in a few schools, include simple simulations, educational games aimed at reasoning and deductive skills, and simple word-processing programs. Some of them are based on cognitive research that specifies learning problems that students have in a subject or the underlying skills that are needed for mastery of complex processes like writing. Others are based simply on the intuitions of skillful practitioners and instructional designers.

An example of a level 2 application is a simulation system, DynaTurtle, developed at the Massachusetts Institute of Technology. This system uses an extension of the programming language LOGO, in which students can give instructions that cause objects on the screen to move in accordance with the principles of Newtonian physics. Development of the system was based on the empirical recognition that students often fail to understand the fundamental concepts of a subject, even when they are able to apply formulas mechanically or to recite factual information. Cognitive research, some done using the DynaTurtle system, has documented the fact that it is often students' preconceptions about a subject that interfere with the acquisition of more accurate conceptual understandings, in spite of considerable formal instruction.

For example, students typically believe that an object in motion has some force that was transmitted to it by whatever caused the motion and that is dissipated as the motion continues. This experience-based "mental model" prevents most students from understanding momentum. Systems such as DynaTurtle take advantage of dynamic computer graphics to provide concrete information about general principles in response to student actions, enabling them to see the consequences of their ideas and to develop experience-based understandings of abstract concepts.

Assessment

Level 2 applications are being developed by educational researchers, often in collaboration with teachers and instructional designers, and frequently implemented by commercial organizations. While these instructional systems represent significant advances in educational software design, there are a number of ways in which they do not tap the full potential of information technologies to improve education:

- The systems often address isolated components of the curriculum and isolated instructional goals, and are thus difficult for teachers to integrate into classroom activities. Further, there are many learning topics for which an adequate cognitive analysis has not been performed yet, and little is known about how such specialized systems affect students' understanding of the instruction that is normally given on the same topics. Continued applied research is needed to address these concerns.
- Many of the systems have goals that are not thoroughly understood by teachers. Games addressing general cognitive skills are likely to puzzle those teachers who have a background in programmed instruction. A system for correcting a misconception in physics will seem mysterious to a teacher who shares the misconception, as many do.
- The systems that have been developed are based on ideas about helping students learn, rather than helping teachers teach. The orientation is a reasonable one, but is not optimal for encouraging enthusiastic adoption by teachers. Research and development focused on the use of computer-based materials in relation to the performance of teaching, either as a direct means of achieving teachers' goals or as supplementary materials that are viewed as beneficial, could increase the effective use of educational technology.

Level 2 applications represent a diversity of efforts that is important for discovering ways to use technology effectively for education. Besides their instructional value, such systems provide important artifacts around which to build continued research into the specific learning mechanisms that enable them to be effective. Most importantly, the involvement of teachers and instructional designers in this development makes it easier to begin implementing technological and other changes in the schools in an orderly way.

LEVEL 3: ADVANCED THEORY AND SYSTEMS DEVELOPMENT

Advanced systems combining symbolic computational capabilities, artificial intelligence, and cognitive theory have begun to be developed in a few well-equipped laboratories. As with level 2, each system addresses a relatively specific aspect of skill or knowledge but the systems differ in that they are informed by an explicit model of human learning and expertise. In addition to their eventual practical value as components of integrated learning systems, these systems and the computational equipment on which they are developed provide a powerful set of learning laboratories around which to deepen our knowledge of how conceptual change and understanding occur and of the methods by which we can bring them about more reliably.

Building on groundwork by behavioral scientists and interdisciplinary work in linguistics, psychology, and computer science, developments in cognitive science have coalesced around a theoretical model of the human mind as a complex information-processing system in which information is organized, stored, and operated on in ways that we are only now beginning to characterize in detail. Within this framework, learning may be viewed as a process of information transfer, leading to useful lines of inquiry concerning the structure of information to be transferred, the characteristics of the receiver (including the biases and constraints of the "system architecture"), and a theory of the transfer process itself.

At the same time, methods in artificial intelligence for constructing expert systems have developed rapidly. Two features of this work are important for education. First, expert systems in a given subject are useful components of computerized tutoring systems; their commercial development, therefore, represents a potentially valuable resource for education. More importantly, the construction of an expert system requires an explicit representation of the knowledge that constitutes expertise. This work, along with empirical studies of human expert performance, has been a means of gaining crucial insights into the nature of human knowledge, skilled problem-solving, and reasoning.

Expert Tutoring Systems

Early efforts to adapt expert systems to education foundered because the organization of factual and procedural knowledge within the systems bore little resemblance to human thinking and reasoning. Further work informed by cognitive research has succeeded in specifying more completely the types of knowledge required for mastery of problem-solving in a given subject, and the results have been applied to education and training. For example, developers of NEOMYCIN refined, restructured, and semantically rationalized the knowledge base of the seminal expert system MYCIN to teach medical diagnosis to students.

The panel believes that continued efforts to formalize the conceptual knowledge, problem-solving strategies, and inferential processes of experts will enable the creation of more robust systems capable of solving novel problems in a given topic or domain. Such systems should be able to tackle unanticipated, student-posed problems and, perhaps more importantly, to show students the underlying processes by which the systems choose among alternative solutions.

Expert Diagnostic Systems

Detailed specifications of the knowledge and skills required for work in a discipline make possible the construction of expert diagnostic systems that are able to isolate underlying causes of student errors. Construction of such systems requires a complete theory and symbolic representation of all the components of an overall skill. However, it also requires (a) a theory of how those subskills are mislearned that accounts for observed student mistakes; (b) a theory accounting for student modification—e.g., shortcuts—of formally taught procedures; and (c) a model of the "noise" in the students' execution of the skill.

One such system is DEBUGGY, an expert system for diagnosing procedural errors or "bugs" in base-10 subtraction. Starting from its data base of more than 130 empirically determined "primitive" or single bugs, DEBUGGY synthesizes a procedural model of the exact nature and cause of a student's deviation from a correct subtraction procedure. What makes the development of such a system especially noteworthy for education is its robustness: its ability to handle compounded student errors that arise when students have more than one bug, as well as their careless mistakes that can disguise systematic errors. Tested with over 4,000 students, DEBUGGY is able, using a set of search heuristics, to distinguish among 10^8 possible hypotheses about the nature of a student's errors.

Besides possible direct use in the classroom, computer-based diagnosis is an important part of developing cognitively based testing methods that measure a student's partial knowledge and pinpoint his misconceptions. The information provided by DEBUGGY can be used to provide personalized and precise remediation not hitherto possible.

Such systems also have been adapted for use in training student teachers, focusing their attention on the underlying causes of student learning problems and on strategies that they themselves can use in diagnosing students.

Perhaps most importantly, diagnostic systems provide a new tool for validating precise theories of learning. Previously unobserved knowledge states can now be deduced, enabling learning theorists to postulate changes in mental states and to verify them on the basis of successful prediction about observable behavior.

Expert Coaching Systems

The successful construction of expert systems and diagnostic systems in various domains has enabled the construction of computer-based coaches. Existing computer-based coaches such as WEST and WUMPUS construct a differential model of the student's conceptual strengths and weaknesses both through error diagnosis and by comparing student behavior with what the computerized domain expert would do at any given point in solving a problem. However, effective coaching also requires knowledge about how and when to intervene. Although such knowledge is usually held tacitly, even by master teachers, existing computer-based coaching systems have begun to characterize this kind of knowledge as a collection of rules that govern computer intervention.

Future work will be directed toward meeting two major challenges. The first one is to expand the current existence proofs for expert coaching systems to include a wider variety of domains, more flexible and psychologically informed tutoring rules, and more robust domain expert systems that mimic closely the reasoning processes of human experts. Such systems will provide a computational workbench for refining our understanding of tutoring knowledge and a building block for commercial educational software developers.

Secondly, systems must be both cost-effective to construct and readily modifiable by end-users. The emergence in artificial intelligence of new programming tools makes it possible to reduce the production time for a domain-specific coach from years to months. In addition, use of these sophisticated programming environments can enable researchers to make the system's tutoring knowledge more accessible and modifiable by teachers and educational technologists.

Discovery Learning Environments

Systems are also under development employing artificial intelligence techniques and based on cognitive research that does not require direct computer intervention into the learning process. Instead, these systems provide a rich environment that aids students in discovering how to solve problems more effectively by recording, structuring, and "playing back" the student's own problem-solving processes.

Such systems are based on cognitive research that shows that successful learners exhibit a high degree of reflective skill, reviewing their progress toward a goal, analyzing their misunderstandings, and revising their strategies accordingly. Poorer learners tend simply to skip over or invent temporary solutions to aspects of a problem that they do not understand.

Computational systems have been developed in algebra and geometry that handle low-level computational operations, thus enabling students to focus attention on their strategic choice as they solve problems. The systems then provide a structured "trace" of their solutions so that students can see the alternative paths that they have tried. Preliminary tests with students show that they are motivated to review their choices and to formulate hypotheses leading to more effective choices in later problems.

The capability of information technologies to record and abstract students' problem-solving processes provides unique leverage for researchers interested in the role of error in learning, while providing students with a mental "mirror" of their own thinking. Armed with these tools, researchers can determine empirically the effects of increased self-reflection and of extended work and revision on fewer problems—particularly for poorer learners.

Such learning environments are impractical with the technology currently used in schools. But technology is already available in research laboratories that could make apprenticeship and experiential learning possible on personal computers of the late 1980s, along with methodologies to develop the scientific understanding of learning needed to design the computational environments and to formulate effective learning tasks.

The technology will be here; the need is for an adequate scientific understanding of how to exploit it. The next section describes five specific research problems on which the field is ready to make significant advances.

RESEARCH NEEDS

CONCEPTUAL CHANGE

Current educational practice appears to be remarkably unsuccessful at connecting school knowledge with experiential knowledge. A major research task is to redesign the educational materials, taking into account the deeply held beliefs and conceptions that students have when they begin instruction. Characterizations, already begun, of students' naive conceptions are not enough to enable remediation. A more adequate understanding of how change in the organization and representation of knowledge occurs during learning is needed.

Students' initial conceptual framework also includes powerful conceptual categories, called "phenomenological primitives," by which they organize their understanding and learning. Growing understanding of the function and origins of these structures can improve our ability to craft educational experiences and restructure domain knowledge so that students can integrate newly acquired knowledge into the informal knowledge and reasoning that they already have, while preparing for the deeper conceptual changes necessary for correct understanding.

CHANGES IN SUBJECT MATTER CONTENT

Widespread use of computation is leading to substantial changes in both the form and the content of procedural skills that it is important to teach to students. Obvious examples arise in mathematics, where procedures for calculation that are difficult to learn can be performed routinely on inexpensive calculators. Somewhat less obviously, many procedures that are taught in mathematics are done to facilitate hand calculations. For example, many procedures for "simplification" in elementary algebra produce expressions that are simpler if one has to evaluate them on paper, but are by no means simpler conceptually or for operations with a digital computer.

In addition, a major reexamination of the curriculum is called for as a result of new opportunities for integrating the content of education both within and between subject-matter disciplines. For example, the capabilities exist for creating educational data bases that can manifest connections between parts of the curriculum. Such connections are rarely made explicit. For example, Ohm's law is usually taught in elementary physics without reference to its being an example of a simple proportional function in mathematics. Design of curriculum sequences with relations like this in mind could increase significantly the coherence of the educational experience and the knowledge that results from it. Cross-referenced data bases also would aid teachers, who could generate easily examples of concepts in their instruction that relate to other areas of their students' work.

MEASUREMENT AND EVALUATION

A critical goal of research on learning is the development and testing of methods for measuring deep cognitive outcomes to guide the design of instructional systems. Current diagnostic techniques enable us to isolate misunderstandings in the acquisition of procedural skills at a new level of detail; however, the way in which more general misunderstandings develop and how they affect subsequent learning is less well understood. Research is needed to assess students' grasp of the conceptual bases and purposes of these procedural skills that they learn to perform, and to develop tests of their ability to modify a procedure flexibly when required.

The ability of sophisticated computational systems to record and abstract the actions of a student while engaging in various complex activities provides a rich source of data for learning researchers. The analysis of verbal protocols—that is, stream-of-consciousness records of subjects' thinking while performing complex cognitive tasks—has led to many advances in cognitive science. With information technologies enabling the automatic collection of electronic protocols and providing inferencing tools to aid in analyzing student thought processes, we are likely to achieve a new set of advances.

MOTIVATION

In addition to cognitive factors, intrinsic motivation is important to successful learning. One counter-intuitive result of recent research indicates that extrinsic reward—for example, an extra play period—can reduce the motivation of students who already engage in a particular activity willingly.

While motivation theory has traditionally been a "messy" and difficult area of scientific inquiry, the use of electronic learning environments as motivational laboratories, in which variables and effects can be isolated and manipulated subtly, has had a significant impact on research already. For example, study of the motivational factors of computer games—level of challenge and user control, completeness of information and its effect on curiosity, and the role of fantasy—establishes a beginning taxonomy of variables that can be examined within the context of experimentally designed learning environments. The popularity of fast-action video games and their possible consequences for the cognitive development of American youth necessitate a better understanding of how the slower, more difficult mastery of cognitive skills can be made rewarding for a broad range of students.

Research is needed also to assess the implications of what appear to be the motivating effects of the computer as a medium, independent of instructional content, and to determine its impact on the transfer of motivation and learning to nonelectronic settings. Work also is needed to determine the interactions between factors such as aptitude and a given motivational treatment within the electronic learning context and to experiment with feedback and reward structures.

IMPLEMENTATION ISSUES

As stated earlier, many past efforts to reform education have had only limited success. Analyses of educational change have identified a set of basic pedagogical techniques and curricular contents that have been particularly resistant to change. Cognitive analysis suggests that, in many cases, these are precisely what must change if improvements in the quality of learning for all students are to take place. How to achieve these changes is therefore a crucial research issue.

Appropriate sorts of demonstration projects can help to spread awareness among educators about the types of learning activities, motivational effects, and other qualitative outcomes that are possible with flexible electronic learning systems. Work at centers focusing on the use of educatinal technology in schools also should contribute to our knowledge about how to effect institutional change.

Curricular and pedagogical change in the direction of individualized discovery learning and greater integration among subjects are impeded by a lack of objective measures of the benefits of those forms of instruction. With the development of validatable measurements for deep cognitive outcomes and demonstrations through testing of the positive effects of cognitively based electronic learning, it may be possible to generate a discerning market demand for fundamentally new kinds of learning environments. Such an approach to needed reform has the clear advantage of mobilizing market forces, creating an accountability structure for the educational software industry, and preserving local choice in formulating school policy.

New learning systems themselves can become the carriers of cognitive theory and new principles of pedagogy into classrooms and homes. Teachers will then be able to assume a freer role in aiding and assisting students in their individual efforts. Further, teachers can use technology to improve their teaching techniques. Cognitively based learning systems can be used simultaneously to teach professional courses in education, to engage teachers in the same kinds of activity-based learning experiences that they will use with students later, and as objects of study about underlying cognitive processes.

Information technologies can have a profound impact on the sociology of the classroom. Several pilot studies have shown that the use of computers as tools and as communications devices tends to alter relationships between students and teachers and among peers. The design of collaborative activities around learning systems can help engineer a new classroom infrastructure within which innovations are less disruptive.

CONCLUSIONS

The time is ripe to develop a new scientific discipline as a fundamentally new resource for American education. A multiyear effort focused on advanced basic research using information technology to enhance and study learning could provide a major synthesis of the knowledge, theory, and methodology that now exist in the relatively separate disciplines of artificial intelligence, cognitive psychology, linguistics, and educational research in school subjects. Integral to this research synthesis will be experimentation with and construction of advanced computer-based learning environments. Such environments will serve as laboratories for deep cognitive research, while also providing significantly new educational resources directed at real educational problems.

By developing new learning environments and evaluating their pedagogical effectiveness, gains will be made in understanding the mechanisms of experiential learning, the integration of formal and informal knowledge, the integration of conceptual knowledge and procedural skill in a domain, and the integration of knowledge across present subject matter disciplines. The technological and theoretical products of this work will provide models of learning systems that can increase the number of American students who succeed in acquiring the knowledge and cognitive skills necessary for effective work and citizenship.

RESOURCES REQUIRED

In order to make these advances, increased resources of three kinds are needed, beyond the levels that are currently available in the field:

1. Sophisticated Symbolic Computation
Wider access to sophisticated computational equipment of the kind now used in standard artificial intelligence research is required for effective contributions to the research and development effort. At present, the availability of such equipment to cognitive and educational researchers is extremely limited, and well beyond funding allocations in these areas. The computational resources needed for cost-effective research and development exceed those required for the delivery of educational systems in schools, if they are to support the interactive programming methodology that makes possible rapid exploratory and experimental software development. Therefore, we recommend a substantial increment of funding to make sophisticated interactive computation and associated networks much more widely available than is now the case.

2. Interdisciplinary Research
Research on learning and on the development of educational systems development requires the talents of computer scientists, cognitive scientists, artificial intelligence researchers, social scientists, and subject matter experts. At present, opportunities for these sorts of interdisciplinary partnerships are limited and occur primarily as isolated projects in universities. A systematic effort to establish an integrated discipline, composed of individuals trained in the relatively new theoretical and technical skills needed to contribute effectively to the growth of knowledge about learning and its applications, is necessary. The creation of one or two special research centers (perhaps modeled after the Institute for Theoretical Physics at Santa Barbara—and similarly established with support by the National Science Foundation or other federal agencies) would provide powerful interdisciplinary resources to further the timely growth of the scientific knowledge requisite to significant educational applications of advanced technology.

3. Stable Funding and Talent

Existing programs represent a useful range of perspectives for near-term research and development, and it seems appropriate to continue with that mixture of support. However, current funding levels reflect a very low priority for basic research. Recent oscillations in funding levels at the National Science Foundation (Science Education Directorate) and at the National Institute of Education have been catastrophic. Stable funding for basic research and training is essential if this field is to attract first-rate scientists and graduate students. Programs are needed to support postdoctoral fellowships in which individuals trained in one of the relevant disciplines could visit for extended periods research institutions where they could obtain training and experience in the other fields.

GENERAL COMMENTS

The levels of computer hardware and instructional programs now prevalent in schools, while probably beneficial, do not add sufficient value to ordinary instruction to warrant increased federal investment to promote the distribution. The more fruitful use of federal funds for distribution of technology is to continue to encourage development and integration into schools of level 2 and level 3 systems. The existing programs are needed to identify specific instructional difficulties and opportunities for improvement and to increase the use of instructional systems in schools and other educational settings.

The advanced research and development efforts suggested here are intended to complement programs that are already in progress or planned at federal agencies. Certainly, existing efforts will be enhanced significantly by the increased understanding of learning, epistemology, and advanced symbolic computation that would result from more advanced projects.

We are, therefore, suggesting a clear recognition of (1) the need to capitalize sophisticated symbolic computation equipment for use both as research tools and as hosts for experimental learning environments; (2) the interdisciplinary nature of basic research in this area; and (3) the need to attract first-rate scientists and graduate students to the field. Significant increments in research capabilities can be achieved with relatively modest amounts of additional funding for equipment and to support interdisciplinary research and postdoctoral fellowships. The results could bring about major advances in our ability to use information technologies in precollege education.

RESEARCH BRIEFING PANEL ON CHEMICAL AND PROCESS ENGINEERING FOR BIOTECHNOLOGY

Arthur E. Humphrey ***(Chairman),*** *Lehigh University, Bethlehem, Pa.*

James E. Bailey, *California Institute of Technology, Pasadena, Calif.*

Harvey W. Blanch, *University of California at Berkeley, Berkeley, Calif.*

Stuart E. Builder, *Genentech, Inc., San Francisco, Calif.*

Robert R. Dorsch, *E. I. du Pont de Nemours & Company, Inc., Wilmington, Del.*

Wolf Hanisch, *Cetus Corporation, Emeryville, Calif.*

James Lago, *Merck & Company, Inc., Rahway, N.J.*

Alan S. Michaels, Alan Sherman Michaels, *Sc.D., Inc., New York, N.Y.*

John A. Quinn, *University of Pennsylvania, Philadelphia, Pa.*

Kevin Ulmer, *Genex Corporation, Rockville, Md.*

Daniel I. C. Wang, *Massachusetts Institute of Technology, Cambridge, Mass.*

Neal R. Amundson ***(Ex-Officio),*** *University of Houston, Houston, Tex.*

Staff

Robert M. Simon, ***Project Director***

Robert M. Joyce, ***Rapporteur***

Robin Illsley, ***Administrative Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON CHEMICAL AND PROCESS ENGINEERING FOR BIOTECHNOLOGY

EXECUTIVE SUMMARY

The phenomenal progress in molecular biology, genetics, and biochemistry over the past two decades now makes possible the programming of living cells of virtually every type—microbial, plant, and animal—to generate products ranging from simple molecules to complex proteins. The United States leads the world in research in this "new" biology. Maintaining that leadership and commercializing the fruits of that research are vitally important to the future economic well-being of the nation and to its international technological stature. A tremendous number of opportunities await commercial exploitation. This will not occur rapidly without adequate knowledge of the fundamental engineering principles underlying bioprocess scale-up. What is needed is a knowledge base in process engineering that combines the skills of the biologist and the chemical engineer in order to bring these opportunities to fruition. The rapid establishment of this knowledge base will require the cooperation of academia, industry, and government.

The intense international competition in commercializing high-technology opportunities, particularly in biotechnology, highlights the need for the United States to maintain its scientific and industrial leadership. Despite this nation's preeminent position in basic biological science, our foreign competitors have established commercial positions by practicing effective and forward-looking biochemical engineering. They can be expected to be no less aggressive as the results of U.S. research in the "new" biology diffuse abroad.

Estimates of the potential worldwide markets for biologically derived products fall in the range $40–$100 billion annually by the year 2000. First entry into these markets will be critically important, and major shares of the worldwide bioproducts market will be captured by those countries who possess the needed national capability in engineering research and personnel.

There are not enough biochemical engineers in the United States today, nor are there enough faculty to train the engineers who will be needed by the emerging biotechnology industry. Fewer than 20 U.S. Departments of Chemical Engineering have meaningful biochemical engineering programs; collectively, these are training fewer than 60 Ph.D. and M.S. graduates annually. This panel estimates that the annual need for graduate-level biochemical engineers over the next decade will average 2–3 times that number. Moreover, the training of biochemical engineers should expose them to the power of modern biological science in manipulating and controlling cellular biosynthesis, and to the utility of life sciences techniques and methods in solving critical large-scale bioprocess problems. Prompt and effective exploitation of the "new" biology is critically dependent on the improvement of this life science/biochemical engineering interface.

The technology needed to capitalize on the discoveries in modern biology falls into two areas:

- Increasing our biochemical engineering expertise in the design, scale-up, and optimal control of large-scale processes and culture of bacterial, plant, and animal cells.
- Expanding our knowledge base so that we will be able to achieve the large-scale recovery and purification of complex, unstable biological macromolecules, as well as of simple biologically derived organic compounds (e.g., ethanol, amino acids) from the dilute and impure solutions that are the initial products of biosynthesis.

Our recommendations for government action are aimed at developing this needed knowledge base in biotechnology and at reducing the communications gap between life scientists and biochemical engineers. Specifically, we recommend that appropriate federal agencies:

- Provide funding at adequate levels for basic research in biochemical and process engineering.
- Establish grants for cooperative, cross-disciplinary research that will stimulate the development of research groups led by joint principal investigators from both the life sciences and biochemical engineering.
- Provide funds for academic units to acquire, operate, and maintain the special types of equipment needed for biochemical engineering research.
- Establish faculty development awards to allow tenured faculty to broaden their expertise in biotechnology through special study and leaves, and to provide promising young faculty with the opportunity to immerse themselves in biotechnology research by temporarily relieving them of some of their teaching commitments.
- Establish institutional grants to train graduate students, thus providing the personnel that will be needed if U.S. industry is to capitalize on emerging opportunities in biotechnology.

INTRODUCTION

The phenomenal progress in molecular biology, genetics, and biochemistry over the past two decades now makes possible the programming of living cells of virtually every type—microbial, plant, and animal—to generate products ranging from simple molecules to complex proteins. The United States leads the world in research in this "new" biology. Maintaining that leadership and commercializing the fruits of that research are vitally important to the future economic well-being of the nation and to its international technological stature. The utilization of living cells and their enzymes for the industrial manufacture of useful products—for which the term "biotechnology" has been coined—requires the adaptation of these systems to large-scale production facilities. This transformation from laboratory experimentation to commercial production is the province of the chemical engineer, or more particularly the biochemical engineer.

Biochemical engineering translates the knowledge developed in biological and chemical science into practical industrial processing; the biochemical engineer bears the same relationship to the life scientist that the chemical process engineer does to the chemist. Chemical engineering enriched by active involvement in the "new" biology can provide a knowledge base to meet the challenges that will be faced by the next generation of chemical engineers in commercializing biotechnology. The establishment of this knowledge base will require the cooperation of academia, government, and industry. Chemistry and chemical engineering have enjoyed a long tradition of active cooperation to the benefit of the chemical industry. However, the interface between chemical engineering and biology needed for practical implementation of biotechnology has not yet evolved. What is needed is a sound curriculum that will promote this cooperation and train the biochemical engineers who can develop the knowledge base required to serve the biotechnology industry.

OPPORTUNITY ASSESSMENT

The opportunities for commercialization of the products emerging from the "new" biology are both tantalizing and highly diverse; among the more exciting prospects are applications falling into the following five categories:

Human and Animal Health Care. A revolutionary new family of diagnostic products based on enzymes, monoclonal antibodies, and other genetically engineered proteins promises to provide quick and highly accurate detection of immunity to or infection by viral and bacterial diseases, of susceptibility to autoimmune diseases, of the presence of genetic defects, or of the existence of neoplasms. Other significant opportunities include novel prophylactic products, exemplified by vaccines for the prevention of viral, bacterial, and protozoal diseases such as hepatitis, typhus, and malaria; new therapeutic biologicals for the treatment of cardiovascular and cerebrovascular disease, neurological and CNS diseases, rheumatoid arthritis, diabetes, and cancer; peptide hormonal substances that increase milk production of dairy cattle and stimulate growth, fecundity, and enhanced feed utilization of cattle and other farm animals.

Human and Animal Nutrition. Applications include the utilization of low-cost carbon sources for new microbiological and enzymatic syntheses of amino acids, sugars, and edible fats and oils, and the use of large-scale fermentation of low-value feedstocks in the manufacture of nutritionally balanced single-cell protein for human or animal consumption.

Agricultural Chemicals and Related Products. Prospects for applications of biotechnology include biologically derived fungicides and herbicides that are highly potent, highly specific, and environmentally safe; plant growth regulators to stimulate crop productivity, reduce plant sensitivity to environmental stress, and reduce fertilizer requirements; new techniques for crop propagation and strain improvement that employ genetically manipulated plant cell clones to replace seed.

Environmental Improvement and Protection. New microbial and enzymatic techniques for removing or destroying toxic pollutants in municipal and industrial wastes provide a promising horizon for the application of biotechnology to environmental problems.

Natural Resource Utilization. New microbiological processes for the recovery of metals and nonmetals (e.g., iodine) from low-grade ores and subsurface aquifers are under development, promising to increase substantially our supply of essential minerals. The possibility of increasing recovery of petroleum from depleted reservoirs by microbiological means is being explored, as are *in situ* microbial and enzymatic techniques for transforming solid fossil carbon sources (coal, shale) into gaseous and liquid fuels.

These tremendous opportunities cannot be exploited without adequate knowledge of the fundamental engineering principles underlying bioprocess scale-up. We face critical engineering problems in the design and control of large-scale cell-culture processes for the manufacture of new bioproducts, and of processes for the efficient large-scale recovery and purification of bioproducts from the complex mixtures in which they are made.

INTERNATIONAL COMPETITION IN RESEARCH

Who will lead the commercialization of the "new" biology? U.S. progress in basic biochemical engineering research has lagged substantially behind its progress in the life sciences. The United States has also lagged behind Western Europe and Japan in support for basic chemical engineering research. This situation must be corrected if our national ability to lead the commercialization of the "new" biology is not to be seriously compromised. Otherwise, the Japanese and the Europeans will certainly take the leading role. The following case histories are illustrative.

- The U.S. consumer was spared much of the world sugar crisis through bioreactor technology based on Japanese patents. This technology uses a continuous, immobilized-enzyme process that converts the glucose derived from domestic cornstarch to high-fructose corn syrup, a commercial sweetener. The process, with $1.5 billion annual sales, now uses $50 million annually of immobilized enzyme purchased from European companies.

- U.S. companies led the world in manufacture of amino acids prior to the 1960s. Then, an intensive Japanese effort began to improve amino acid production through biological and biochemical engineering techniques. First, classical genetic strain improvement techniques were applied to isolate mutant microorganisms that produce greatly increased quantities of amino acids. Second, biochemical engineers utilized very large bioreactors and developed efficient methods for recovering the product. Today, Japan dominates a business with annual world sales of $1.7 billion.
- Process technology for penicillin manufacture was developed in the United States in the 1940s. Protected by patents, several U.S. companies dominated the world penicillin market for many years. Now that the patents have expired and U.S. companies have decided to invest in research on new, higher-return products rather than on process improvement, European companies with modern process technology account for the preponderance of worldwide sales of natural penicillin.

Despite the preeminent U.S. position in basic biological science, our foreign competitors have established commercial positions by practicing effective and forward-looking biochemical engineering. They can be expected to be no less aggressive as the results of U.S. research in the "new" biology diffuse abroad. Foreign developments in biochemical engineering already attest to that aggressiveness.

- Basic technology for membrane separation of biomolecules was invented in the United States, but the Germans and Japanese are the ones applying it to separations of enzymes and amino acids from complex mixtures.
- Technology for very large (400,000 gal) continuous fermenters was developed and is being practiced in the United Kingdom. This development pushes biochemical engineering to limits not yet explored in the United States.
- Although the use of fermentation to produce ethanol is an ancient technology, more efficient immobilized-cell, continuous processes have been conceived, and Japan has established the first demonstration-scale plant.

A survey of biochemical engineering research and education in the United States, Western Europe, and Japan was carried out by several members of this panel. A striking difference between U.S. and overseas activities is the magnitude and organizational structure of the government-supported research and development effort in biotechnology. West Germany, Japan, and the United Kingdom each have three federally supported institutes dedicated to biotechnology; there are none in the United States. These nine institutes bring together academic and industrial investigators, feature cross-disciplinary activities, and have impressive operating budgets.

In 1983, the West German Gesellschaft für Biotechnologische Forschung (GBF) had an operating budget of DM37.1 million ($14 million), and the Institut für Biotechnologie II in Jülich had DM11.1 million ($4.3 million) for biochemical engineering alone. The last figure nearly equals the entire NSF annual budget in biochemical engineering. A quantitative comparison of space and associated equipment for biochemical engineering shows that West Germany's research space is nearly double that of the United States.

In Japan, the MITI 10-year plans have strong research and development components in biochemical engineering; three of the nine areas emphasized are bioreactors, animal cell culture, and membrane separation. Japanese government support of membrane separation research and development alone amounted to $20 million in 1983. The comparable NSF effort is less than $1 million.

It is obvious that countries such as Germany and Japan are laying a foundation of research and trained personnel as part of their strategy for intense international competition in the arena of biotechnology. The potential economic rewards for success are very great; estimates of the potential worldwide markets for biologically derived products fall in the range $40–$100 billion annually by the year 2000, corresponding to about 15 percent of the estimated total annual market for chemicals. First entry into these markets will be critically important in international competition, and major shares of the worldwide bioproducts market will be captured by those countries who possess the needed national capability in research and personnel.

STATUS OF U.S. RESEARCH AND EDUCATION

There are not enough technologically competent biochemical engineers in the U.S. today, nor are there enough faculty to train the engineers who will be needed by this emerging industry. According to an analysis by members of this panel, there are roughly 80 chemical engineering faculty members in the United States who currently have some involvement in biochemical engineering research and at least 50 more who have indicated an active interest in the field. Fewer than 20 U.S. Departments of Chemical Engineering have meaningful biochemical engineering programs; collectively, these are training fewer than 60 Ph.D. and M.S. graduates annually. Yet this panel estimates that the annual need for graduate-level biochemical engineers over the next decade will average 2–3 times that number. The current personnel demands of industry are recruiting younger faculty members and recent graduates who would otherwise be strengthening the university teaching and research capabilities in this rapidly growing field.

The costs of doing research in some areas of biochemical engineering are increasing rapidly as the field moves away from an Edisonian approach to an emphasis on discovering and applying fundamental engineering principles. Biochemical engineers are beginning to investigate the principles underlying the design and control of reactors for growing mammalian cell cultures. Such work holds great promise in reducing to practice the substantial U.S. lead in monoclonal antibody research and in large-scale cell culture for the production of tissue plasminogen activator for the rescue of heart attack victims. Virtually no university in the United States has cell culture laboratories equipped for engineering studies. Such laboratories are costly. A critical review of what constitutes adequate funding to perform state-of-the-art research should be part of an overall plan for ensuring U.S. competitiveness in this area.

Most important, there is a critical need to strengthen the life science/biochemical engineering interface. The historic success of the development of the U.S. chemical industry has in large measure been due to close collaboration and communication between the chemist and chemical engineer in both academia and industry. This collaboration has been fruitful because both disciplines share a common language and comprehension of physicochemical laws and principles, and a mutual appreciation for the special skills that each brings to bear on the adaptation of chemical discoveries to industrial practice. In contrast, few life scientists are aware of the engineering principles and practical problems associated with the scale-up of biological processes or with the large-scale processing of bioproducts. Conversely, few chemical engineers are sufficiently knowledgeable in the principles of modern molecular and cellular biology, microbiology, genetics, and biochemistry to permit their effective communication with life scientists. This lack of communication inhibits progress on the unique problems that make biological processes particularly difficult to engineer:

- Living organisms can undergo spontaneous mutations, which can have dramatic effects on process operations.
- Processes must be carried out under aseptic conditions to prevent contamination by ubiquitous, undesired organisms.
- Biological processes usually take place in dilute aqueous solution, imposing the need to separate the product from large amounts of water.
- Many of the high-value-in-use products of biotechnology are fragile, hard to purify, and structurally complex.

The training of biochemical engineers should expose them to the power of modern biological science in manipulating and controlling cellular biosynthesis, and to the utility of life sciences techniques and methods in solving critical large-scale bioprocess problems. Prompt and effective exploitation of the "new" biology is critically dependent on the improvement of this life science/biochemical engineering interface; this is one of the most urgent needs confronting biotechnology today.

INTELLECTUAL CHALLENGES AND CRITICAL NEEDS

Because of the rapid progress in the life sciences, commercialization of the fruits of modern biology is not opportunity-limited. The principal limitation is the lack of a technology base that can address the following key challenges:

- Increasing our existing biochemical engineering expertise in techniques of large-scale cell culture not only using simple microorganisms (e.g., bacteria and yeasts) but also utilizing the full range of plant and animal cell systems.
- Expanding our knowledge base so that we will be able to achieve large-scale processing, including the recovery and purification of complex, unstable biological macromolecules (e.g., antibodies, peptide hormones) from product mixtures, as well as the economic and efficient recovery of simple biologically derived organic

compounds (e.g., ethanol, amino acids) from highly dilute and impure solutions.

Specific areas of research aimed at meeting these challenges are as follows:

BIOREACTORS

Mechanically agitated reactors currently used for antibiotic fermentations are often poorly or not at all suited to meet the diverse demands of bioprocessing. New techniques are needed for large-scale culture of plant and animal cells as well as the new microorganisms now being engineered for bioproduct synthesis. Fundamental knowledge of how physical and environmental factors influence intracellular biosynthetic pathways is essential to the development of such techniques. Bioreactor research will require the hybridization of such sciences as molecular and cellular biology, microbiology, and cell physiology with basic engineering skills, chemical kinetics, thermodynamics, fluid dynamics, heat and mass transport, and precise bioprocess control—a rare combination today in either industry or academia.

An important parallel direction of bioreactor research is the development of methods for utilizing free or immobilized enzymes, enzyme combinations, or nongrowing whole cells as catalysts for biosynthesis. The effective utilization of such diverse forms of biocatalysts requires exploration of fluid bed, fixed bed, membrane, cell recycle, tubular, and other reactor types in which to carry out biosynthesis. The challenge is to translate the existing knowledge base for chemical reactors into biosystems, where strict asepsis, complicated biological regulation processes, enzyme and cell fragility, cofactor regeneration of enzyme activity, and maintenance of cell energy provide additional dimensions to the problem. Process fundamentals developed in benchtop systems must be tested in large-scale equipment in order to gain understanding of the scale sensitivity of biosystems.

SEPARATION AND PURIFICATION

The product stream generated in a bioreactor must go through separation and purification operations in order to isolate the desired product in adequate purity. For fragile, high-unit-value products intended for human or animal health applications, there is a premium on processes that minimize product deterioration and maximize purity. For low-unit-value products where competition lies in nonbiological industrial synthesis the premium lies in recovery processes that are energy-efficient and have high recovery efficiency and low environmental impact. The areas of separation and purification science from which biotechnology can benefit through increased R&D effort fall into three principal categories:

Modification of Conventional, Large-Scale Industrial Separation Methods. Such processes as ion exchange chromatography, pressure-driven membrane separation, electrodialysis, and liquid-liquid extraction are now employed industrially to recover or purify antibotics and simple molecules such as citric acid. They are generally too nonselective or too destructive to be useful for processing fragile protein biologicals. Refinement of these techniques is needed to render them suitable for modern bioproduct recovery.

Adaptation of Biochemical Laboratory Separation Methods to Large-Scale Bioprocessing. Life scientists have developed extremely powerful and sophisticated separation tools such as electrophoretic separation and affinity separation that are potentially applicable to large-scale separations. However, much needs to be learned about the molecular mechanisms and kinetics of these processes before a rational approach to scale-up can be initiated.

Novel Separation and Purification Concepts. There is need for discovery and development of novel separation and purification concepts that embody combinations of physical, chemical, and biological phenomena outside the armamentarium of conventional chemical process technology. Several such processes are being studied in Japan and Europe. Some are based solely on physicochemical principles, for example aqueous two-phase separations employing water-soluble synthetic polymers, or multiple-field fractionation. Other process concepts represent a constructive synthesis of biochemistry and cell biology with industrial chemistry and chemical engineering. Examples include separation based on modification of permeability, separation by selective enzymatic transformation, and separation by genetically manipulated intracellular processes.

BIOPROCESS INSTRUMENTATION AND CONTROL

The successful operation of bioreactors and downstream processing equipment requires sophisticated process control, which depends upon accurate measurement of critical process variables and upon the use of advanced estimation algorithms, process models, and control strategies. Current biosensors and control methods often do not provide the desired reliability or the capability for sensitive regulation of process conditions. Furthermore, precise bioprocess control often requires on-line monitoring of the concentration of a complex biological substance for which a specific sensor does not exist. The use of enzymes, monoclonal antibodies, and even whole living cells as components of electrochemical and optical detectors offers promise in solving some of these problems.

Research on bioprocess models is needed to develop optimal control strategies and in order to extract the most useful information from measurements. Formulation of these models will require more advanced control algorithms and a greater knowledge of the effects of engineering parameters on cells and complex molecules.

TRAINED PERSONNEL

Although the need for trained personnel has already been addressed in this report, one aspect of this problem deserves mention in this section on critical needs. Constructive interaction between life scientists and engineers will be an essential element of innovative bioprocess development. The federal government can make an important contribution at this critical stage in the development of the new bioprocess science and the commercialization of biotechnology by stimulating a research environment in the academic community that will bring together life scientists and chemists with chemical engineers to focus upon bioprocessing needs and problems, and to train a generation of engineers who will be capable of designing, building, and operating the biotechnical industry of the future.

CONCLUSIONS AND RECOMMENDATIONS

If the United States is to capitalize on the revolutionary advances in biology, it must develop both a fundamental knowledge base in chemical and process engineering for biotechnology and a cadre of trained personnel so that U.S. industry can translate fundamental research results into commercial products. The infant U.S. biotechnology industry does not yet have a sufficiently broad product base to support a significant amount of fundamental research in academic institutions. In the short term, the federal government has a critical role to play in attaining the following objectives.

- Establish the fundamental knowledge base to support the design, scale-up, and optimal control of reactors and processes for the large-scale growth of microbial, plant, and animal cells required for full exploitation of the new biology.
- Develop the fundamental knowledge base for large-scale separation and purification processes that will be required to produce the spectrum of potential biochemical products from simple organic molecules to complex proteins.
- Train the next generation of biochemical engineers in a research environment that provides cross-disciplinary exchange of knowledge, industrial collaboration, availability of state-of-the-art facilities and equipment, and support levels that will assure research productivity.

The following recommendations can achieve these objectives if they are implemented at a level that will gain a competitive advantage for the United States. They will assure the development of the needed knowledge base and reduction of the communications gap between life scientists and biochemical engineers. Accordingly, we recommend that the appropriate federal agencies:

- Provide funding at adequate levels for basic research in biochemical and process engineering that will address the research areas of critical importance in this report (see Appendix).
- Establish grants for cooperative, cross-disciplinary research that will stimulate the development of research groups led by joint principal investigators from both the life sciences and biochemical engineering.
- Provide funds for academic units to acquire, operate, and maintain the special types of equipment needed for biochemical engineering research.
- Establish faculty development awards to allow tenured faculty to broaden their expertise in the principles and techniques of biotechnology through special study and leaves, and to provide promising young faculty with the opportunity to immerse themselves in biotechnology research by relieving them from some of their teaching commitments.
- Establish institutional grants to train graduate students, thus providing the personnel that will be needed if U.S. industry is to capitalize on emerging opportunities in biotechnology.

APPENDIX
RESEARCH AREAS OF CRITICAL IMPORTANCE

A. Bioreactors

1. *Quantitative Molecular and Metabolic Reaction Engineering*
   - *In vivo* control mechanisms and their relationship to metabolic regulation
   - Bioenergetics of biochemical cascade reaction networks
   - Chemical regulation in biosynthesis
   - Kinetics and mechanism of action of genetically engineered microorganisms

2. *Fundamental Understanding of the Physical and Environmental Factors in a Bioreactor Needed for Successful Scale-Up*
   - Fluid dynamic behavior of bioreactors and its influence on biological behavior
   - Gas-liquid and liquid-solid mass transfer and heat transfer in bioreactors
   - New reactor systems for viscous fermentations
   - High-pressure, nonmechanically agitated bioreactors
   - Development of perfusion reactor systems that are compatible with large-scale production
   - Integration of bioreactor operation with product recovery
   - Bioreactors for nonaqueous systems to reduce downstream processing costs
   - Containment in bioreactors designed for genetically engineered organisms

3. *Expanding the Knowledge Base on Such Biocatalysts as Immobilized Enzymes or Whole Cells*
   - Mechanistic understanding of enzyme stability and inactivation
   - Techniques for immobilizing enzymes and whole cells for large-scale industrial processes
   - Bioreactor systems for reactions involving cofactor regeneration
   - Transport mechanisms in heterogeneous biocatalysis

4. *Engineering Fundamental for the Design of Reactors for Handling Plant and Animal Cells*
   - Quantitative definition of the stability of, and the effects of

mechanical shear on, genetically engineered animal and plant cells

- Physiological characterization of animal and plant cells
- Characterization and fundamental understanding of surface-cell interactions for those animal cells that require attachment to surfaces in order to grow and produce ("anchorage-dependent" cells)
- Development of surfaces for anchorage-dependent cells that can be used in large bioreactors
- Transport phenomena in plant cells due to formation of large cell aggregates

B. Separation/Purification

1. *The modification of conventional, large-scale, industrial separation methods to render them more efficient, reliable, and/or economical when employed for biological product recovery. Examples in this area include:*
   - Column chromatography: development of new chromatographic materials that do not alter the structural characteristics of biological macromolecules
   - Pressure-driven membrane separation processes: membranes that are less subject to fouling by biological process streams
   - Liquid-liquid extractors: development of liquid two-phase systems that do not alter the structural integrity of biological macromolecules
   - Electrodialysis: development of membranes with selectivities useful for product streams from the "new" biology

2. *Adaptation of Biochemical Laboratory Separation Methods to Large-Scale Bioprocessing*
   - Electrophoresis: Protein behavior in combined electrical and low gravitational fields as well as in centrifugal fields and in flow fields in gradient pore-size gels
   - Affinity chromotography: mechanisms and kinetics of antibody-antigen interactions
   - Affinity chromotography: factors that influence the association/dissociation equilibria in affinity separations

3. *Novel Separation and Purification Process Concepts*
   - Fundamental principles in aqueous two-phase separations

- "Multiple Fields" such as electrical, magnetic and gravitational fractionation processes
- Separation principles and technologies based on selective enzymatic transformations
- Biological and chemical alteration of cell-wall permeability to facilitate product separation
- Integration of modern biology with downstream recovery science to alleviate potential purification problems by involving biochemical engineers in the design of recombinant cells to perform the biosynthesis

C. Bioprocess Instrumentation and Control

1. *Measuring Instruments for Important Biological Variables:*
   - Sensors for viable cell density, biological product concentrations, cellular metabolic state, and mixed culture compositions
   - More reliable and sensitive instruments for measuring concentrations of dissolved oxygen and carbon dioxide, ionic strength, and other abiotic parameters
   - Biosensors based on enzymes, antibodies, and cells with adequate stability and durability

2. *Mathematical Models to Aid in Process Design, Control, and Optimization:*
   - Process descriptions based on key features of biological mechanisms rather than on empirical correlations
   - Models that accurately represent the effects of simultaneous changes in several process variables

3. *Advanced Control Strategies for Regulation of Complex Biological Processes:*
   - Data analysis and interpretation algorithms based on biological structure and statistical filtering principles
   - Control for new multiphase, multifunction bioprocesses (one example would be a bioprocess combining the reaction step with the product separation step)
   - Methods for identifying and counteracting undesired cellular control mechanisms
   - Controls to maximize the useful lifetime of immobilized biocatalysts and separation media

RESEARCH BRIEFING PANEL ON HIGH-PERFORMANCE POLYMER COMPOSITES

David McCall *(Co-Chairman), AT&T Bell Laboratories, Murray Hill, N.J.*
Rudolph Pariser *(Co-Chairman), E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.*
Paul J. Flory, *Stanford University, Stanford, Calif.*
Frank Gerstle, *Sandia Laboratories, Albuquerque, N. Mex.*
Allan S. Hay, *General Electric, Schenectady, N.Y.*
Ted Helminiak, *Wright-Patterson AFB, Ohio*
Frank Kelley, *University of Akron, Akron, Ohio*
William J. MacKnight, *University of Massachusetts, Amherst, Mass.*
Arthur B. Metzner, *University of Delaware, Newark, Del.*
Seymour Newman, *Ford Motor Company, Detroit, Mich.*
Manuel Panar, *E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.*
Darrell H. Reneker, *National Bureau of Standards, Gaithersburg, Md.*
L. E. Scriven, *University of Minnesota, Minneapolis, Minn.*
Leslie Smith, *National Bureau of Standards, Gaithersburg, Md.*
George M. Whitesides, *Harvard University, Cambridge, Mass.*
Dick J. Wilkins, *General Dynamics Corporation, Fort Worth, Tex.*

Staff

William Spindel, ***Study Director***
Robert M. Simon, ***Staff Officer***
Robert M. Joyce, ***Rapporteur***
Robin Illsley, ***Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON HIGH-PERFORMANCE POLYMER COMPOSITES

EXECUTIVE SUMMARY

High-performance polymer composites are formed from exceptionally strong, highly oriented, long fibers, such as graphite, aramid, or glass, which are placed in a desired orientation and bound together by a polymer matrix. The strengths and moduli thus achieved are on a par with those of the strongest available structural metals, but are substantially higher on an equal-weight basis.

The utility of advanced composites is based on their anisotropic physical properties and their unique strength and modulus per unit weight. These properties have led to significant adoptions as structural members and skins of military aircraft. As experience has been gained in the design and fabrication of these materials, they are finding increasing use in space vehicles, civilian aircraft, recreational equipment, and moving parts in textile machinery. Large-scale applications in automobiles, heavy equipment, robotics, and building and bridge construction are foreseen. We are now on the threshold of realizing the vast potential of advanced composites, which promise to play a major role in the U.S. economy and defense.

The field of advanced composites is an international arena. The present U.S. position is strong in chemistry, materials engineering and application, but Japan dominates in carbon fiber technology. In Japan, a broad MITI program covering advanced composites is in place. In West Germany, a new Max Planck Institute for polymer research will impact on composite technology.

Research needs include the following areas:

- New fiber compositions, polymer matrices, and fabrication methods are needed to broaden applications and to reduce costs.
- Understanding of composite structure/property relationships is lacking. The fiber/matrix interphase region has not been well characterized, although it is critical to composite performance.
- Knowledge of failure mechanism is primitive. Flaw identification, growth, and elimination need study. Response to differing load conditions and environmental factors has not been described in fundamental terms. Life prediction, reliability, and testing are critical matters that await characterization and understanding at the molecular level.
- A science of composite design and processing needs to be developed and must involve extensive computer-based modeling for design, engineering, and manufacture. Joining and repair processes must also be developed on a fundamental basis. The toxicity, environmental effects, and eventual disposal of composites must be examined for their consequences on fabrication and use of these materials.
- Reinforcement of a polymer matrix at the molecular level, with stiff, rodlike polymer molecules, has been demonstrated and should be studied to learn how to develop optimum properties in such composites.

The rapid technological advances in high performance composites over the past decade have outpaced the development of the underlying basic science. Research in the area calls for communication and collaboration among several disciplines, including chemistry, physics, chemical engineering, mechanical engineering, and materials science. These groups tend to be insular in present U.S. university structures. The future will belong to those who can bring together the necessary disciplines and construct a coherent intellectual structure that will form the basis for advancing to new levels of understanding.

Research on advanced composites in U.S. universities is embryonic and frequently includes only one of the several science and engineering disciplines needed to advance the subject. A realignment of the traditional academic structure appears necessary if the needed interdisciplinary research is to be effective. Fewer than 30 U.S. universities conduct research on composite materials, involving about 40 equivalent full-time faculty. Multidisciplinary organizations devoted to composites exist at only two of these universities. Moreover, a substantial fraction of this university research is devoted to product design and to lower-performance short-fiber composites.

Estimates of overseas support for composites research indicate that several countries have efforts that are comparable to that of the United States. However, their programs are growing more rapidly than ours and are concentrated on the interdisciplinary approach that is essential to progress.

It is recommended that a modest number—perhaps three, as it becomes feasible to staff them—of interdisciplinary centers devoted to basic research on advanced composites be established in universities. The proposed Engineering Research Centers are a good model, and research on composites should be encouraged in them. In addition, more research specifically concerned with advanced composites should be encouraged in existing academic interdisciplinary centers devoted to materials science. The objectives of these centers would be to develop the scientific underpinning that is essential to advance the national position in advanced composites and to develop a pool of trained scientists and engineers for future personnel needs in research, development, and teaching in this high-technology area. Industrial involvement in the form of intellectual presence and research cooperation in the campus centers is regarded as essential.

INTRODUCTION

Structural composites are not new. Nature has been perfecting these materials for millions of years, resulting in such composite structures as wood, bone, or clam shells. These structures comprise combinations of components specifically chosen in terms of amount, size, shape, orientation, and interface to achieve a set of overall properties that is optimized for its intended purpose.

This is also the central purpose of man-made high-performance composites, a relatively new area of technology that holds promise for revolutionary advances in structural materials. These composite materials are filling crucial national needs. They are already essential to the defense establishment and the space program, and are beginning to have significant impact in important areas of the national economy, particularly aircraft, recreational products, textile machinery, and ground transportation.

Precursors to advanced composites are found in such diverse developments as the automobile tire and short-fiber reinforced plastics. In recent years, several classes of high-performance structural composites have been developed, including multicomponent ceramics, short-fiber reinforced metals, and polymer composites based on strong, highly oriented, continuous fibers developed through advances in inorganic and polymer chemistry. High-performance polymer composites are made by arranging arrays of strong, stiff fibers in a desired orientation, usually in close-packed alignment, and binding them together with polymer matrices. This new class of polymer composites provides combinations of strength and modulus substantially superior to those of structural metals and alloys on an equal weight basis. This report is concerned only with these high-performance polymer composites.

The principal impetus for the development of high-performance polymer composites came from military aircraft and space vehicles, which needed structural components of minimum weight and with high strength in prescribed directions. Further impetus was provided by the adoption of these composites in civilian aircraft to achieve fuel savings, in recreational equipment, and in moving parts of textile machinery to achieve higher operating speeds. We are now only on the threshold of realizing their vast potential. Significant improvements in cost and performance are expected and should greatly expand their uses.

The United States pioneered the use of advanced composites and is currently in a strong position in this technology. However, significant programs on composites research are being implemented in other countries. Specifically, a broadly based academic-industrial program, initiated by MITI, is in place in Japan. Germany's composites research will likely be strengthened soon by a new Max Planck Institute for polymer science and engineering in Mainz.

The key issue addressed in this report is the need for increased research in basic science and engineering in high-performance composites to provide the technical foundation for meeting important national needs. The rapid technological advances in high-performance composites over the past decade have greatly outpaced the development of the underlying basic science. Further improvement in such areas as composite manufacturing and processing will undoubtedly emerge from continuing applied research and development. However, basic research in this area must be strengthened and broadened to acquire needed knowledge and to assure a supply of professional scientists and engineers with the background necessary to advance the technology. An investment in basic research would be a high-leverage investment.

This report is organized into the following sections: current and projected APPLICATIONS of high-performance composites; SCIENCE AND TECHNOLOGY—the present and prospective state of the art, international competition, and goals and directions for research; EDUCATIONAL NEEDS; and CONCLUSIONS AND RECOMMENDATIONS.

APPLICATIONS

The evolution of high-performance composites is founded in the opportunity to build light-weight, high-strength, high-modulus structural parts that can be tailored to meet specific loading conditions. Because these composites can be made with superior strength and modulus in one or two directions, structural parts can be designed specifically to accommodate the anticipated direction of loading, in contrast to isotropic metallic parts. Advanced composites are currently high in cost, reflecting costs of fibers and manufacturing processes; both cost elements are decreasing and should continue to do so.

High-performance composites are widely used in space vehicles because they lower launch weight and have good dimensional stability over a wide range of temperatures owing to their low coefficients of thermal expansion. These composites are used as structural components in many military and civilian aircraft and will comprise a larger fraction of the structural members in the coming generation of aircraft. Range, payload, fuel consumption, tooling costs for parts, and radar profile all enter into design considerations, and the payoff can be enormous. Japan has identified the aircraft industry as one of five in which they expect to be a leader in the 1990s, and their commitment to composites research is part of their strategy.

Sporting goods, including spars for boats, racing car bodies, bicycle frames, tennis racquet frames, fishing rods, and golf club shafts, bring high-performance composites to the public eye. Composites provide superior performance, and users are willing to pay for it.

Use of advanced composites in automobiles and other surface vehicles has been limited, and a number of problems must be solved before this industry can realize the benefits of these materials. An acceptable balance between processing speed and product quality has not yet been achieved. Useful technology for joining and repairing is not in hand, and long-term dimensional stability must be improved. However, there are important benefits to be gained by using composites in such vehicles. Tooling costs are much lower than for steel, providing greater manufacturing flexibility, more rapid design changeover, and reduced capital investment. It has been estimated that the 1,500 body parts in one automobile could be reduced to a substantially lower number by substituting high-performance composites for metals. Composites are less subject to corrosion than metals, and vehicle weight is reduced, with consequent fuel savings. At present, advanced composites are finding applications in leaf springs and drive shafts in production vehicles, and experimental composite engines have been constructed and tested. Automotive applications in the United States could expand dramatically in response to lower composite costs and the development of lower-cost manufacturing processes for using them.

Military uses in addition to aircraft are many and growing. As in sports equipment, performance rather than cost is the driving force. Examples include helmets, body armor, parts for personnel vehicles and for tanks, light-weight segments for movable bridges, carriages and braces for artillery, and portable rocket launchers.

High-performance composites have great promise in industrial machinery as reciprocating, oscillating, or rotating parts where the weight, strength, fatigue life, noise, and natural vibrational frequency of metals limit operating speeds and thus productivity. These composites are being used for moving parts in textile machinery, where higher operating speeds and productivity are paramount. A carbon fiber/epoxy centrifuge rotor can be spun at twice the speed of one made of high-tensile steel because of its high specific strength, providing a separation efficiency four times that of the steel rotor. Significant productivity increases can be foreseen by taking advantage of the high specific strength and modulus, lower moment of inertia, and the self-damping characteristics of composite parts.

High-performance composite parts have potential uses in robots and heavy construction machinery, where the low moment of inertia of a cantilevered member translates into lower energy costs and more accurate control of its positioning. Another potentially large area of use is load-bearing members of buildings and bridges, where the strength/-weight ratio, self-damping, and corrosion resistance offer substantial benefits.

In summary, advanced composites are being used in a number of industries that are vital to the national economy. Their uses in these industries will unquestionably expand with anticipated improvements in cost, manufacturing speed and consistency, and particularly in basic understanding of the critical properties that must be measured to ensure optimum design and prediction of the useful life of structural parts.

SCIENCE AND TECHNOLOGY

FIBERS

High-strength fibers are the primary load-bearing components of high-performance composites; they comprise about 60 percent of the volume of the composite and largely control both properties and cost. Carbon, aramid, and glass are the most widely used fibers. They have tensile strength/density and modulus/density ratios several times that of steel and represent a new technology that is just beginning to demonstrate its potential.

Glass fibers were the basis for some of the earliest reinforced polymers. The development of composites of continuous glass fibers, although largely empirical, led to the first controlled processing conditions and micro-structure that took advantage of the fiber strength.

Carbon fibers are stronger, stiffer, and more costly than glass. They are produced by pyrolyzing oriented polyacrylonitrile to leave carbon, which is graphitized. The process may include hot stretching. Lower-performance carbon fibers are made from coal tar or petroleum pitch. Empirical adjustment of processing parameters has resulted in commercial carbon fibers in which either tensile strength or modulus is optimized. The strain of carbon fibers at break is about 1 percent.

Aramid fibers are prepared by spinning a polymer that has a highly rigid linear molecular structure from an oriented anisotropic phase, in which the molecules are spontaneously oriented over microscopic dimensions. During spinning the oriented regions are aligned with the fiber axis to give high strength and rigidity. Because of the low density of the polymer (1.44), the specific tensile strength of the fiber is more than 5 times that of steel. The elongation at break is greater than that of nonmetallic fibers of similar modulus, and this property can be an advantage in composite design because toughness is related to energy absorption before break. Aramid fibers have excellent thermal

TABLE 1
PROPERTIES OF FIBERS

		Steel	Aluminum	E-Glass	S-Glass	High-Strength Graphite	High-Modulus Graphite	Aramid	Boron
Tensile strength (10^3 psi)	Achieved 1984	500	100	350	650	720	350	525	500
	Theoretical	3,000	1,000	1,000	1,280	6,000	6,000	4,000	5,800
Tensile modulus (10^6 psi)	Achieved 1984	30	10	10	12.8	36	75	18	58
	Theoretical	30	10	—	—	100+	100+	29	—
Density (g/cm^3)		7.8	2.7	2.54	2.5	1.75	1.8	1.44	2.7
Thermal expansion ($10^{-6}/°c$)		11	24	5	3	−1	−1	−2	5
Specific strength (10^6 in.)		1.8	1.0	3.9	7.3	11.5	5.4	10.5	5.2
Specific modulus (10^8 in.)		1.1	1.0	1.1	1.4	5.8	11.7	3.5	6.0

properties (although inferior to those of inorganics), good chemical stability, and low flammability. Like carbon, they can be made in a range of properties that optimize either strength or modulus.

The spinning of an anisotropic phase has been extended to both solutions and melts of newer polymers. For example, experimental samples of polymers such as polybenzothiazole appear to have better properties for composite use than the commercial aramids. The spinning of anisotropic melts offers the potential of lower costs than spinning from solution, but the polymers that can be spun from melts tend to give fibers with lower use temperatures. Another example is polyethylene, which usually has low modulus and strength. It has been fabricated into ultra-high-strength fibers by gel spinning. Processes for making fibers from poly(ethylene terephthalate) have recently been patented by a Japanese group.

Boron fibers are prepared by decomposition of a volatile boron compound on a resistively heated fine tungsten filament, which is converted into tungsten boride and serves as the fiber core. Fibers of silicon carbide can be made by a similar process or by pyrolysis of a polycarbosilane fiber. Fibers of silicon nitride and alumina have also been developed.

Since the various fibers have different mechanical properties as well as costs, it is not uncommon to create hybrid composite structures comprising more than one kind of fiber to balance performance/cost requirements.

Tensile failure in carbon, glass, and other inorganic fibers is typically initiated by a flaw or surface growth step that causes local stress concentration and produces a simple fracture perpendicular to the fiber axis. The weak lateral bonding and oriented polycrystalline nature of organic fibers such as aramid produce a fibrillar failure. Considerable splitting along the fiber axis occurs and results in a mixed shear and tensile failure.

In compression, carbon fibers usually fail by combined interface separation and microbuckling. Aramid develops shear or kink bands that form at relatively low stresses but do not produce catastrophic failure. Glass and boron fibers fail in compression when an axially oriented flaw is activated by expansion in the noncompressed direction. Compressive strengths of boron and glass composites, as well as those made from alumina and silicon carbide, exceed their tensile values.

The origin, growth, and interaction of fiber flaws with the composite matrix are poorly understood. Failure mode is not predictable from tensile strength and modulus, and the question of what measurable properties are critical to behavior in use has not been resolved. It is recognized that the time-dependent behavior of fibers under stress is important. Most test numbers are derived from short-term static tests, which do not give a reliable indication of long-term life in use.

The strength and modulus of current carbon, aramid, and glass fibers are significantly below the calculated theoretical values. However, substantial improvements in the properties of all three have been made over the last 10 years; while it is unlikely that the theoretical properties will be attained, there is good reason to expect that improvements will continue to be made. Moreover, the possibilities in organic polymers for new high-strength, high-modulus fibers have only begun to be explored, and it is expected that organic fibers with strengths 2–3 times those currently available will be made.

MATRICES

The matrix component of a composite serves primarily to bind the load-bearing fibers. The matrix is the minor component of the composite, and, while it does not usually support stress, flaws or damage in it may lead to failure of the composite. The properties of the matrix usually determine the limits of temperature and chemical environment that the composite can withstand.

Current polymer matrix materials are principally thermosetting epoxy resins, which are chemically cured to produce highly cross-linked structures. The curing process requires elevated temperatures and substantial times. Cross-linked epoxies have very low fracture toughness, and their composites are intolerant of damage because cracks will propagate readily through the matrix. Tougher materials have been developed by incorporating a rubbery dispersed phase in the epoxy, but these polymers have lower continuous use temperatures and lower modulus. Epoxy composites can be used at temperatures up to about 175°C. However, under hot/wet conditions they lose properties at lower temperatures. New high-temperature/high-modulus resins, such as polyimides, bismaleimides, and acetylene-terminated oligomers based on high-temperature resins, promise improved matrix properties with lower moisture sensitivity and faster curing. These matrices offer the potential of use temperatures in the range 200–370°C. However, their toughness is still low, and they must also be cured for significant times at elevated temperatures.

Tough, high-temperature-resistant and solvent-resistant thermoplastics with low moisture sensitivity are being explored as matrix resins. Examples are polyamide-imides, polyphenylene sulfide, polyetherimides, and polyether-etherketone (PEEK). They can be processed in the melt or in solvents that must subsequently be evaporated. They are strong candidates for replacing thermosets because of their shorter fabrication cycle, which does not require a chemical cure. Property improvements should result from their greater inherent toughness and the fact that some of them are less water-sensitive than thermosets. Thermoplastics creep under stress at elevated temperatures, and current materials have use temperatures in the range 150–230°C.

The ideal resin for a matrix material would be a low-viscosity material that could easily be combined with the reinforcing fiber and could then be converted rapidly and reproducibly at low temperatures into the final composite structure. Important properties of the matrix include high and/or low temperature capability, resistance to water and solvents, toughness, and a coefficient of thermal expansion that matches that of the fiber. Although it is unlikely that all of these properties will be found in one matrix material, the potential of new polymer-forming monomers and reactions, copolymers, block copolymers, and blends should lead to a variety of high-performance matrices.

COMPOSITES

One method of manufacturing high-performance composites involves the resin-impregnation of bundled continuous filaments of a reinforcing fiber to form a continuous tape or fabric. This prepreg is subsequently plied in layers that may have the same or different alignments. The plied structure is then subjected to pressure and heat, which yields a consolidated composite structure for subsequent assembly or manufacturing steps. The combination of structural design efficiency and the flexibility of the manufacturing process through the layering of directional plies provides the desired structural properties.

Another manufacturing process involves winding a coated fiber onto a mandrel. A computer-controlled payout system capable of multiaxis precision winding can generate virtually any geometry on an irregularly shaped mandrel. The fiber-wound structure is impregnated with polymer, which is then cured, and the composite structure removed from the mandrel.

Although some aspects of composite technology are quite advanced, in others our scientific understanding and processing methods are clearly inferior to, for example, the knowledge of structure-property relationships and fabrication techniques for steel or aluminum alloys.

Since structural metals are the principal competition for composites, increased acceptance of high-performance composites depends on their becoming as reliable, simple, and economical to produce as highly developed alloys.

High-performance composites have presented engineers and materials scientists with a variety of new challenges. Analytical and numerical techniques had to be developed to treat anisotropic mechanics problems. In particular, linear elastic mechanics needed for orthotropic plates and shells, and the micromechanics for calculating homogeneous, thermomechanical properties are now in hand. Techniques also exist for predicting thermoelastic properties and for detecting mechanical or compositional flaws for the more common composite components. Processing and fabrication of thin laminates into consistently high-quality materials have become routine.

However, composite manufacturers have not yet developed techniques for consistent production of thick laminates for highly stressed structural elements; fibers often move during the fabrication process, with deterioration in composite properties. Producing a void-free, high-fiber-volume laminate still involves much empiricism, since the process of creating and curing a laminate is difficult to model. Aspects of high-performance composites that are less developed, but also needed for design, are data on dynamic tensile and fatigue strength. Progress has been made in predicting the effect of flaws on failure modes dominated by the resin, but not those dominated by the fiber.

COMPETITIVE WORLD POSITION

The United States has a strong position in aramid fibers and a reasonable level of activity in developing other organic fibers with comparable or better properties. However, aramid fibers may be soon commercialized in both Europe and Japan, and the activity on other organic fibers appears to be at least as strong in Japan. MITI has designated the development of a "third-generation" fiber as a government-subsidized project, beginning in 1983 and targeted for practical applications in 5 years. Because of the overwhelming importance of the load-bearing fibers in composites, the field is sensitive to breakthroughs in stronger fibers. Current patent activity suggests that such a breakthrough is as likely to come from Japan as from this country. For example, processes to make high-strength fibers from poly(ethylene terephthalate) have recently been patented by a Japanese group.

Much of the technology used for manufacture of carbon fibers in the United States is licensed from Japanese companies. Although there is research activity in this country, some companies have chosen joint ventures with Japanese firms to take advantage of Japan's developed technology. The high level of Japanese carbon fiber technology also suggests that they may produce many of the expected future advances in these materials.

The United States currently has a strong position in composite manufacturing and processing technology and leads the world in developing major applications in such industries as aircraft, sporting goods, and automotive. However, there is growing overseas activity in composites technology. Foreign manufacturers are gaining valuable experience by subcontracting for the manufacture of large structural parts for new-generation U.S. commercial aircraft, using technology supplied by the United States. European countries are as active as the United States in putting high-performance composite parts into their military aircraft. Japanese and West European technologies are beginning to rival ours.

RESEARCH DIRECTIONS

The rapid technological development of advanced composites has outpaced the basic understanding of the structure and performance of these materials. A strengthening of basic research in this area is needed to provide a firm scientific foundation for the anticipated future advances in composite technology. General directions for such research are summarized below.

We need better understanding of molecular structure/property relationships of fibers and of how such properties translate into their behavior in a composite. Composites are subjected to long-term cyclical loading, and we need to understand the causes of fiber fatigue under dynamic conditions. Although there is some knowledge of how different fibers fail under tensile and compressive loads, there is essentially no knowledge of how such failure may be dictated by the fiber microstructure. We also need to investigate the effects on composite properties of fiber molecular structure and macrostructure: molecular composition, molecular orientation, surface structure, size, and geometrical placement in the composite, and how composite properties are affected by processing.

We need to define the physical and chemical parameters of matrix resins that can be translated into prediction of performance in a composite. The curing of thermosetting matrices should be studied to determine the effects of curing conditions on the cross-linked network structure, physical properties, and fiber interactions with the cured resin. Structural modifications of epoxies and other thermosets that will increase their toughness should be explored. Thermoplastic polymers have potential advantages over thermosets in offering greater toughness and requiring shorter composite fabrication times. The interaction of thermoplastic polymers with fibers at the interface will be substantially different from that of thermosets and need to be understood at the molecular and morphological levels. Research may point the way to reducing the tendency of thermoplastics to creep at elevated temperatures under load.

There is much to be learned about the fundamentals of the fiber-polymer interface: How does adhesion affect the load bearing, and how can optimum adhesion be achieved; what are the molecular structure and mobility of the polymer at the interface; what can be learned about nonequilibrium polymer conformations at the interface; how does the interface affect the kinetics of growing polymer chains; what properties can be achieved with various polymers and fibers when the interface is prepared under ideal laboratory conditions?

We lack the knowledge base for combining fibers and matrices to take maximum advantage of their properties. Nondestructive test methods must be developed to determine whether the manufacturing process has attained the desired microscopic and macroscopic structures, with the fibers in their desired positions. We need more precise techniques for predicting composite properties from properties of components, for detecting and predicting effects of flaws, and for predicting remaining service life of a part in use. Because of our limited basic knowledge, composite structures are now designed with considerably greater margins of safety than those used for metal structures, thereby losing some of the inherent superiority of composites. We need better techniques for joining or fastening composite parts to metals with minimum adverse effects on composite strength.

Many potential applications for advanced composites will depend on achieving substantial improvements over current manufacturing speeds and product quality. We must develop automated robotic manufacturing processes to reduce product variability. Composite samples prepared under carefully controlled laboratory conditions should provide benchmark properties against which to assess manufactured parts.

The toxicity of composite components, their long-term environmental effects, and their recyclability need to be studied, and should be considered at the development stage of a composite.

Finally, we should draw upon knowledge of the structure of natural composites such as bone, wood, and clam shells for clues to how man-made composites can be made to approach more nearly their theoretical physical properties.

EDUCATIONAL NEEDS

The rapid technological advances in high-performance composites over the past decade have outpaced the development of the underlying basic science. Research in the area calls for communication and collaboration among several disciplines, including chemistry, physics, chemical engineering, mechanical engineering, and material science. These groups tend to be insular in present U.S. university structures. The future will belong to those who can bring together the necessary disciplines and construct a coherent intellectual structure that will form the basis for understanding present achievements and advancing to new levels.

Research on advanced composites in U.S. universities is embryonic and frequently includes only one of the several science and engineering disciplines needed to advance the subject. A realignment of the traditional academic structure appears necessary if the needed interdisciplinary research is to be effective. Fewer than 30 U.S. universities conduct research on composite materials, involving about 40 full-time equivalent faculty. Multidisciplinary organizations devoted to composites exist at only two of these universities. Moreover, a substantial fraction of this university research is devoted to product design and to lower-performance short-fiber composites. Consequently, too few scientists and engineers are being trained with the broad background needed to advance composite technology, and the gap between supply and demand for trained people is increasing with the rapid growth of the composites industry.

Estimates of overseas support for composites research indicate that several countries have efforts that are comparable to that of the United States. However, their programs are growing more rapidly than ours and appear to be concentrated on the interdisciplinary approach that is essential to progress.

CONCLUSIONS AND RECOMMENDATIONS

Our conclusions are as follows:

- High-performance composites will find increasing numbers of applications in both industry and defense. They are essential to many high-technology industries and they will meet important national needs.
- The present U.S. position in high-performance composites is strong in the areas of chemistry, materials engineering, and applications. There is a strong challenge to the United States in this area from both Japan and Europe.
- Basic research on high-performance composites in the United States must be strengthened to develop the understanding needed to advance the field and to maintain the U.S. position against foreign competition.

It is recommended that a modest number—perhaps three, as it becomes feasible to staff them—of interdisciplinary centers devoted to basic research on advanced composites be established in universities. The proposed Engineering Research Centers are a good model, and research on composites should be encouraged in them. In addition, more research specifically concerned with advanced composites should be encouraged in existing academic interdisciplinary centers devoted to materials science.

The objectives of these centers would be to develop the scientific underpinning that is essential to advance the national position in advanced composites and to develop a pool of trained scientists and engineers for future personnel needs in research, development, and teaching in this high-technology area. Industrial involvement in the form of intellectual presence and research cooperation in the campus centers is regarded as essential.

RESEARCH BRIEFING PANEL ON THE BIOLOGY OF ONCOGENES

Robert A. Weinberg ***(Chairman),*** *Associate Professor of Biology, Massachusetts Institute of Technology, Whitehead Institute, Cambridge, Mass.*

Raymond Erikson, *American Cancer Society Professor of Cellular and Developmental Biology, Biological Laboratories, Harvard University, Cambridge, Mass.*

Fred Rapp, ***Professor and Chairman,*** *Department of Microbiology, The Milton S. Hershey Medical Center, Pennsylvania State University, Hershey, Pa.*

Michael Sporn, ***Chief,*** *Laboratory of Chemoprevention, National Cancer Institute, National Institutes of Health, Bethesda, Md.*

George Todaro, ***Scientific Director,*** *Oncogene, Seattle, Wash.*

George Vande Woude, ***Director,*** *LBI Basic Research Program, NCI Frederick Cancer Research Facility, Frederick, Md.*

Michael Wigler, ***Senior Staff Scientist,*** *Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.*

Rapporteur

Luis F. Parada, *Massachusetts Institute of Technology, Whitehead Institute, Cambridge, Mass.*

Staff

Barbara Filner, ***Director,*** *Division of Health Sciences Policy, Institute of Medicine*

Barbara Mandula, ***Consultant***

Naomi Hudson, ***Administrative Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON THE BIOLOGY OF ONCOGENES

ONCOGENES AND THE MOLECULAR BASIS OF CANCER

INTRODUCTION

Researchers have sought to understand the causes of cancer for many decades. The advances prior to the early 1970s were largely descriptive. However, over the last 10 years, new ideas have been formulated and better technologies developed to address the problem. Possibly the most important result of these labors has been the recent exciting discovery of cellular oncogenes, or cancer genes. Cancer no longer appears as a vaguely defined collection of phenomena with no central mechanistic basis. Researchers are now beginning to define in precise terms—down to the last molecule—the kinds of changes that occur within a normal cell to propel it forward to becoming a tumor cell.

Various agents from outside the organism, such as radiation, chemical substances, and certain types of viruses, have been identified as involved in cancer induction. The accumulated evidence now suggests that viruses and radiation are involved in causing only a minority of human cancers. The suggestion that chemical substances (including agents in tobacco smoke and in the diet) induce the bulk of human cancers did not itself shed light on the molecular processes involved in the transformation of normal cells into tumor cells. A central question remained: How does a cancer-causing chemical alter a cell so that this cell and its descendants exhibit cancer traits? It appeared that the genetic blueprint of these cells, carried in their DNA, must be permanently changed (mutated). This implicated DNA molecules as the central target of carcinogenesis.

The chasm between the formulation of this explanation and a demonstration of its correctness could be bridged only by recent advances in cellular and molecular biology. These advances, both conceptual and technical, led to the discovery of cellular oncogenes and some of the molecular mechanisms of carcinogenesis. Modern DNA technology, developed in studies seemingly unrelated to cancer, has been mustered to isolate specific DNA sequences, analyze them in great detail, and begin to relate their properties to the process of carcinogenesis.

THE SEARCH FOR ONCOGENES

Retrovirus-Associated Oncogenes

Our initial knowledge of cellular oncogenes stemmed from tumor virology. RNA tumor viruses ("retroviruses") were known to cause cancer in lower animals. These viruses harbor cancer-causing genes, called viral oncogenes, that are not genuine viral genes, but rather genes from animal cells that were captured by the viruses in the course of infecting cells. These acquired, previously normal, cellular genes fell under the control of the viral genetic machinery. The result was that a normal cellular gene, a "proto-oncogene," was converted into a potent oncogene. There are more than 20 different oncogenes now known to be associated with a variety of retroviruses, each pointing to the existence of an antecedent proto-oncogene residing among the genes of normal cells.

The normal human genome has precise counterparts (homologues) to the proto-oncogenes originally found in animal genomes. Proto-oncogenes also have been detected in such other organisms as yeast, fruit flies, and mice. The conservation of these proto-oncogenes during evolution suggests that they play a crucial role in cellular and organismic survival.

Discovery of Human Oncogenes

Gene transfer or "transfection" techniques make it possible to extract genes from one cell and introduce them into another. Transfection experiments showed that DNA segments from human cancer cells contain genetic information able to transform normal recipient cells into tumor cells. These DNA molecules contained specific transforming segments termed, once again, oncogenes. Such oncogenes were found in a wide variety of human tumors having no known viral etiology. All the isolated oncogenes were very similar to genes found in normal cells. This confirmed the suspicion that *oncogenes of human tumor cells arise by alteration of genes present in normal human cells.*

Subsequent results were surprising in two ways. First, the human oncogenes were closely related in structure to the oncogenes that have been acquired and activated by cancer-causing animal retroviruses. Thus, mechanisms of cancer causation in humans and animals apparently have common molecular features. By extension, all types of cellular proto-oncogenes that had been activated by animal retroviruses could potentially also be activated in human cancers.

The second surprise occurred when researchers first compared the molecular sequence of a human cancer gene, isolated by recombinant DNA technology, with that of the corresponding proto-oncogene that resides in a normal cell. They found that this oncogene, present in a human urinary bladder carcinoma, had become activated by the mutation of just one nucleotide among the 5,000 nucleotides comprising the normal proto-oncogene. This change resulted in one different amino acid in the protein encoded by that gene. Thus, a very limited change in the structure of a normal protein resulted in a new mandate for the cell that harbored it. That first evidence—now much expanded—showed that important steps in carcinogenesis can be traced to specific, well-defined genetic lesions. Presumably, the mutation responsible for the activation of this particular proto-oncogene occurred in bladder tissue during the course of tumor formation.

About 15 to 20 percent of all human tumors, from a wide variety of body sites, have been shown to carry oncogenes in their DNA. Most of these oncogenes are members of the *ras* gene family. (Oncogenes are given three-letter names, such as *ras, src, myc, most; ras* stands for rat sarcoma, the tumor in which it was first found.) Some of the *ras* genes have been activated in experimental animals with x-radiation and with chemicals known to be mutagenic and carcinogenic. Transfection assays continue to unveil new human oncogenes, at least six of which are now being characterized in laboratories around the world.

Oncogenes and Chromosomal Abnormalities in Malignancies
Many types of human cancers carry specific chromosomal rearrangements. The role of these rearrangements in cancer development was not understood until molecular analysis of human proto-oncogenes became possible.

The best examples of specific chromosomal rearrangements are the "translocations" found in human Burkitt's lymphoma and in mouse plasmacytomas. In both of these diseases, a chromosomal rearrangement relocates the normal *myc* proto-oncogene from its usual chromosomal locus into the vicinity of antibody-producing genes. As a result of this chromosomal translocation, the previously normal *myc* gene is no longer regulated by its usual control mechanisms. With loss of regulation, the *myc* gene is converted into an active oncogene that contributes to the creation of a tumor cell.

In addition to translocations, cancer cells often carry genes in amplified copy number (many copies of the same gene). DNA analysis of some of these amplified genes shows that they are oncogenes. The N-*myc* oncogene is amplified in advanced stages of human neuroblastoma; the *ras* genes have been detected in amplified form in several human cancers; and the *myc* oncogene is amplified in promyelocytic leukemia, colon carcinomas, and small cell lung carcinomas. The amplified copies of these genes cause overproduction of gene-associated products.

BIOCHEMICAL CHARACTERIZATION OF ONCOGENE PROTEINS

Oncogenes exert their effect on cells by means of the proteins whose production is specified by the nucleotide sequences of the genes. The amino acid sequences of almost all the known oncogene proteins have been determined. In many cases, the oncogene proteins closely resemble normal cellular proteins. This close relationship suggests that the oncogene proteins function similarly to normal cellular proteins, but in a deregulated or inappropriate way.

The best characterized of the oncogene proteins is that specified by the *src* oncogene. This protein, associated with the inner surface of the cell membrane, acts as a protein kinase; it attaches phosphate groups covalently to other proteins or to itself. These phosphate groups are linked to the tyrosine amino acids of the target proteins. This specificity distinguishes oncogene kinases from other protein kinases of the cell, which attach phosphate groups to the serine or threonine moieties of their targets. Recently, the *src* protein and a related protein have been found to phosphorylate certain membrane phospholipids. Proteins specified by the *ras* oncogene group have a contrasting behavior, in that they seem to be able to bind or degrade guanosine nucleotides.

Oncogenes and Growth Factors
Recent work has shown that the gene product of one retroviral oncogene, designated *sis*, is related to a subunit of platelet-derived growth factor, a protein actively involved in tissue repair. Similarly the gene product of a second retroviral oncogene, *erb*-B, represents a fragment of the cellular receptor for epidermal growth factor. These findings not only have great import for identifying the nature and mechanism of action of certain oncogenes, but also have brought together two exciting fields of study: oncogenes and growth factors. The work suggests that the intercellular signals for regulating growth (the growth factors) are intimately connected with intracellular mechanisms of growth regulation (oncogene proteins).

MAJOR PROBLEMS IN THE FIELD

HOW ONCOGENES WORK

Having learned to detect oncogenes and their encoded proteins, researchers now are faced with the more difficult task of discovering how individual oncogenes work to cause cancer. It is assumed that the proteins encoded by oncogenes affect the normal cellular pathways that govern growth. How do they do this?

The *src* and *ras* proteins are representatives of the large class of oncogene proteins that are found in the cytoplasm of the cell or near the inner surface of the cell membrane. Other proteins, specified by the *myb* and *myc* oncogenes, are found in the nucleus, where they may control expression of certain cellular genes. The enzymatic functions of these various proteins can provide clues to their physiological roles. Of related interest are cellular regulatory processes in which simple actions of enzymes can result in large signal amplification and produce diverse responses in the cell. Presumably, each of the oncogene proteins will eventually be found to have an enzymatic or regulatory activity that allows it to perturb a cellular growth-regulating pathway. But for the moment, we are ignorant of the nature of these growth-regulating circuits and pathways and of the mechanisms by which relatively simple proteins can elicit the multiplicity of traits associated with cancer cells. This problem of oncogene protein function represents the major challenge of the field for the coming decade.

Interaction of Oncogenes with the Cell Cycle

The simplest view of the growth status of the cell would portray two states—growing and nongrowing. In reality, a complex series of events occurs when a cell emerges from a quiescent state and enters a state of exponential growth. This latter state is characterized by a well-ordered series of events, termed the "cell cycle," that includes chromosomal replication and mitosis. It is probable that some of the complex regulatory decisions required during the cell cycle involve proto-oncogenes. For example, a rapid increase in expression of the normal *myc* gene is one of the initial events that occurs during emergence from quiescence.

Cellular Differentiation

Cancer is a disorder of differentiation as well as of cell proliferation. Stem cells in many tissues form a pool of proliferating cells, some of whose progeny leave the pool and evolve (i.e., differentiate) into specialized cells that constitute the functioning tissue. Stem cells display immortality; they can, in principle, multiply without limitation throughout the lifetime of the organism. Their differentiated daughter cells, however, lose their ability to divide indefinitely. One mechanism by which cancer cells can achieve unlimited replicative potential may depend on inhibiting the ability of these cells to differentiate, trapping them in the pool of replicating stem cells.

One possible role of oncogenes is to maintain the tumor cell in an undifferentiated state, protecting it from progression to the differentiated state with concomitant loss of replicative ability. The process of differentiation and its regulation by (proto-) oncogenes represent major problems in cell biology.

Which Cancer Cell Phenotypes Are Induced by Oncogenes?

The most commonly used definition of an oncogene implies an ability, on the part of a gene, to confer unrestrained, abnormal growth capacities on previously normal cells. Such a definition would suggest that many of the genes involved in the control of normal cell proliferation have the potential, following appropriate activation, to become oncogenes.

However, cancer is more than a problem of unrestrained growth. Cancer cells acquire a number of other traits during their long progression from normality to malignancy. One important trait concerns the ability of these cells to evade being killed by components of the immune system that are responsible for detecting and eliminating abnormal cells that may arise in the body. A second attribute of some cancer cells, metastasis, confers an ability of these cells to invade adjacent tissues and to develop into secondary tumors at distant sites.

Little is known about the biochemical or molecular biologic mechanisms governing these important tumor cell traits. It is not even clear that specific activated genes are responsible for mediating metastasis or immune evasion. We suspect that such genes exist, and that they may properly be included under the rubric of oncogenes, because they probably make important contributions to tumorigenesis. A major problem concerns the nature of these genes and how they are able to produce the variety of traits associated with cancer cells.

PARTICIPATION OF ONCOGENES IN THE MULTIPLE STEPS OF CARCINOGENESIS

Carcinogenesis is believed to involve a series of independent, sequentially occurring steps that enable a cell and its descendants to evolve toward malignant growth. Certain carcinogens, termed "initiators," trigger the process of tumorigenesis. Others, called "promoters," facilitate the development of "initiated" cells into full-blown tumor cells. The effects of initiators and promoters on oncogenes are poorly understood. Several of the early steps may involve activation of various growth-promoting oncogenes. Other alterations may propel the cell toward invasion of surrounding tissue and spread to distant sites.

In one very specific case, that of rat embryo fibroblasts, at least two independently activated oncogenes are required to convert normal cells into tumor cells. Although the generality of this observation remains to be demonstrated, the existing evidence underscores the potential complexity, at the molecular level, of the tumorigenic process. The nature and number of molecular events that transpire in the creation of most human tumors are unknown.

How Many Potential Oncogenes Are Harbored by the Cellular Genome?

The cellular genome carries 50,000 or more distinct genes. Of these, only a small portion are concerned with growth control and with other traits expressed by cancer cells. The number of known proto-oncogenes found in the cellular genome now exceeds 20. These have been discovered by virtue of their association with various retroviruses, or by transfection, or by use of cloned DNA probes derived from the first two approaches. It is likely that additional oncogenes remain to be discovered, perhaps by use of new detection procedures.

How Are Oncogenes Activated From Normal Proto-Oncogenes?
An important question concerns the multiplicity of molecular mechanisms that can mediate the activation of proto-oncogenes into their oncogenic versions. For example, as already discussed, the *ras* oncogenes can be activated by discrete point mutations affecting single nucleotides. Amplified versions of *ras* genes are also found in some human malignancies. Another activated oncogene, the *erb*-B gene, results from the truncation of the gene that encodes epidermal growth factor receptor. As a third example, the *myc* gene becomes activated following chromosomal translocation. The precise mechanisms or agents (such as chemical mutagens) that may lead to these changes must be explored. Although much progress has been made in studying environmental carcinogens and their mode of activation within cells, it is still not clear precisely how many of these activated compounds are able to create oncogenes or induce human cancers.

What Might Reverse or Inhibit Oncogenes?
The panel believes that oncogenes lie at the heart of the cancer process and that they represent the centrally acting agents of cellular transformation, rather than peripheral secondary or tertiary consequences of this process. Much oncogene research has been directed toward understanding the causative mechanisms of cancer, but the discovery of these genes also provokes interest in eventual development of anticancer therapeutics. Reversal of the action of oncogenes or their encoded proteins may allow a reconversion of tumor cells to normal cells. Because certain oncogene proteins are structurally different from their normal counterparts, agents might be developed that specifically recognize and antagonize these proteins.

A major challenge is the development of pharmacologic antagonists of oncogene action. Such antagonists would provide an alternative to the cytotoxic agents currently used in cancer therapy. An alternative therapy might depend upon immunologic reagents (e.g., antibodies), which would recognize and destroy cells carrying oncogene proteins. Unfortunately, such an approach may be extremely difficult because many oncogene proteins are located deep within the cell and thus are hidden from immune mechanisms that are able to survey only the cell surface.

An important, relatively unexplored, possibility concerns the natural mechanisms that the organism employs to check cellular growth. Just as there are growth-stimulating factors, there may be corresponding growth-retarding factors. The expression of proto-oncogenes may normally be counterbalanced by growth-antagonizing "anti-oncogenes," which would be valuable in cancer prevention or treatment.

What Types of Genes Confer Hereditary Susceptibility to Cancer, and How Are These Genes Related to Oncogenes?

Many people have inborn susceptibilities to certain types of cancer. In some cases, the genetic factors conferring susceptibility exert their effects by depriving the individual of normal immune defenses for protecting DNA from accumulating large numbers of mutations. But in most cases, the susceptibility is less general, predisposing to specific types of cancer.

The bases for these susceptibilities are not known, but their genetic components are different from the oncogenes described here. The susceptibility genes are associated with an individual from the time of fertilization of the egg; oncogenes are not genetically transmitted in egg or sperm but develop instead in specific tissues during an individual's lifetime. Several of the susceptibility genes have been associated with specific chromosomal aberrations. The identity of these genes and the nature of their functions are obscure at present.

POTENTIAL AREAS OF RAPID DISCOVERY

GROWTH FACTORS AND SECOND MESSENGERS

The growth of most normal cells in the body seems dependent on growth factors, hormones that are transmitted from cell to cell and provide the external stimuli that trigger growth. Cancer cells often can grow in the absence of the normally required exogenous growth factors. From this perspective, oncogenes can be viewed as agents that confer on cells a growth pattern that is independent of external growth factors.

For example, one oncogene has been found to cause the cell itself to make inappropriately large amounts of a growth factor so that the cell's growth becomes independent of exogenous factors. Another oncogene encodes the structure of an altered growth factor receptor in the cell; this altered receptor may make the cell behave as if growth factors are present even when they are not.

These data and hypotheses have resulted in a merging of two previously separate fields of cancer cell biology, oncogenes and growth factors. An understanding of the mechanism of action of growth factors and their receptors will be central to understanding the physiology of the cancer cell.

Ion Channels and Second Messengers

Regulation of cell growth has often been studied by depriving normal cells of growth stimulatory factors and then inducing them to grow by exposure to these factors. Within seconds of exposure to growth stimulatory factors, a variety of changes are seen in the cell. To cite some examples, Na^+ and H^+ are exchanged across the outer membrane of the cell, the intracellular pH rises, free calcium in the cell rises dramatically, and several of the phospholipids of the membrane are degraded to new compounds.

These phenomena illustrate the activity of so-called "second messengers"—compounds in the cell whose changing concentrations transmit important signals from one cellular constituent to many others. It is highly likely that oncogene proteins and their normal versions modulate the concentrations of these second messengers, which then convey growth regulatory information to various parts of the cell. (It may be that some of the ion-regulating mechanisms that originally evolved for mediating cellular growth control were subsequently adapted for modulating ionic states of nerve cells during conductance of excitatory signals.)

These various considerations indicate that two areas of rapid development over the next several years involve study of intracellular second messengers and study of proteins that regulate movement of ions across membranes.

BIOCHEMICAL AND STRUCTURAL ANALYSIS OF ONCOGENE PROTEINS

Study of oncogene proteins has been hampered because only small amounts of material were available. Successful techniques for cloning the oncogenes and expressing their proteins in bacteria have made increased amounts of these proteins available for study. It soon will be possible to conduct structural analyses of these proteins by means of directed modification of the proteins themselves and x-ray crystallographic analysis. Moreover, the recently developed techniques of directed DNA mutagenesis (whereby specific nucleotides can be altered) will make it possible to alter oncogene sequences and thereby redesign specific portions of the proteins they encode.

Use of synthetic oligopeptide immunogens and monoclonal antibodies will enormously facilitate the localization of these proteins to specific cellular sites, the identification of other proteins with which these proteins interact, and the characterization of complex, multicomponent reactions in which these proteins participate.

GENETIC ANALYSIS IN YEAST AND FRUIT FLIES

Genetic analysis in well-studied lower organisms provides a powerful means of looking at properties of proto-oncogenes and oncogenes and their role in cellular and organismic physiology.

In yeast, *Saccharomyces,* and to a lesser extent in the fruit fly, *Drosophila melanogaster,* it is possible to inactivate a proto-oncogene genetically and to study subsequent effects on the organism's viability, growth, and development. Identification of other genes that interact with oncogenes will provide a new dimension to the current understanding of cellular pathways that are perturbed by oncogene action. Such genetic manipulation is impossible with mammalian cells. Implicit in these experiments is the notion that the molecular mechanisms governing cellular replication in yeast and flies are similar to those in mammalian cells.

The oncogenes discussed until now encode protein products that induce the cell to proliferate. The action of these oncogenes may be balanced by other genes that function to inhibit growth. If such antigrowth genes exist, inactivation of these genes would physiologically mimic the action of oncogenes that actively promote growth.

Some data suggest that certain genes that confer a hereditary, familial predisposition to cancer may act by failing to inhibit growth. Thus, individuals born with an inadequate complement of these "antigrowth" genes may have a lifelong risk of increased cancer incidence. How can we look for these "antigrowth" genes or, more precisely, for the absence of them? It should be possible to use genetics to identify growth-restraining genes in yeast and to employ these yeast genes to find counterparts in mammalian cells.

A further use of genetics derives from our newly gained ability to introduce genes into the germ lines of *Drosophila* and mice. Studies in which oncogenes are introduced into these germ lines will allow researchers to look at the effect that these genes have on normal development of the embryo or on specific organs within the embryo. Such studies promise to provide important new insights into the roles of (proto-) oncogenes in the development of complex organisms and the interaction of certain oncogenes with specific tissues.

THE SEARCH FOR NEW ONCOGENES

Discovery of additional oncogenes is limited by the techniques currently used to detect them. New assays for oncogene activity will be developed over the next several years and presumably will reveal oncogenes whose presence in human DNA was not apparent using existing techniques.

Certain types of oncogenes should be discovered by virtue of their ability to transform indicator cells other than the currently used NIH3T3 mouse fibroblasts. Other oncogenes may be found following direct manipulation of the normal human DNA. Yet others may be found by virtue of their association with specific chromosomal rearrangements.

CANCER DIAGNOSIS

Several reagents soon will be available to detect oncogenes and distinguish their products from those of normal counterpart genes. These reagents may allow earlier identification and greater specificity in detecting tumors than is currently possible.

The success of such diagnosis is predicated upon an issue as yet unresolved: How often are altered oncogenes involved in the causation of human tumors? About 20 percent of human tumor DNAs yield active oncogenes upon transfection into NIH3T3 cells. However, this assay has limitations. In fact, we already know that some oncogenes such as *myc,* which was identified because of its involvement in chromosomal translocations, do not affect NIH3T3 cells and would have been overlooked in such an assay. In other cases, activation of a proto-oncogene may be a necessary but not sufficient step in carcinogenesis and may not warrant a diagnosis of an actively growing cancer.

As methods are improved for discovering new oncogenes, it is the panel's opinion that all human tumors eventually will be found to carry cellular oncogenes. Because oncogenes are important determinants of cellular phenotype, we expect that oncogene diagnosis will provide important prognostic indicators for the clinician dealing with specific tumors. Oncogene diagnosis also may permit better delineation among cancer types and provide better clues to external triggering agents.

NEED FOR CONTINUED BASIC RESEARCH

The great amount of progress in oncogene studies has been possible because of the basic biologic research that preceded and led to the discovery of oncogenes. Continued study in fields such as membrane biology, signal transduction, cellular differentiation, molecular genetics, and cellular immunology will come to bear on the problem of cellular transformation and tumorigenesis. Conversely, the lessons learned from oncogene research will have important implications for many of the unsolved problems of modern biology.

Although the study of oncogenes cannot be expected to resolve all the diverse phenomena that constitute the biology of cancer, we believe that the progress to date has already placed us close to a clear understanding of many of its central mechanisms. Further progress will depend on continuing support of the basic biologic disciplines whose successes have proved so useful in confronting a multitude of biologic problems, cancer being only one of them.

SUMMARY

Within the past 5 years, molecular and cell biologists have made great advances in identifying the molecular mechanisms of carcinogenesis. They have found specific genes within tumor cells, termed "oncogenes," which are responsible for creating many of the abnormal traits of cancer cells. These discoveries provide the foundation for a complete understanding of the molecular basis of cancer within the coming decades.

While these genes have been assigned important roles in carcinogenesis, their precise mechanism of action remains unclear. Much work will be pursued on the structure and enzymatic functions of the oncogene proteins. Thus, certain oncogenes may affect the levels of important compounds in the cell that regulate growth. Others may regulate the expression of certain cellular genes. Yet others may regulate the expression of secreted growth-regulating hormones or receptors used by the cell to detect these hormones.

Future advances in understanding oncogene function will depend on exploiting a variety of experimental tools made available by such diverse basic biologic disciplines as pharmacology, molecular biology, genetics, and cell biology. The presence of related genes in organisms such as fruit flies and yeast will make it possible to study the oncogenes in the context of these genetically simpler organisms. Future advances also will focus on how chemical carcinogens can activate oncogenes and how other agents may antagonize oncogene function and the growth of the cancer cell. New genes may be found that function to control oncogenes and inhibit their functions. Such advances may lead over the next decades to insights into novel therapies that interfere specifically with the growth of cancer cells.

RESEARCH BRIEFING PANEL ON INTERACTIONS BETWEEN BLOOD AND BLOOD VESSELS (INCLUDING THE BIOLOGY OF ATHEROSCLEROSIS)

Alfred P. Fishman ***(Chairman), Director,*** *Cardiovascular-Pulmonary Division, Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pa.*

Ramzi S. Cotran, ***Professor of Pathology,*** *Harvard Medical School, and Pathologist-in-Chief, Brigham and Women's Hospital, Boston, Mass.*

Harriet P. Dustan, ***Director,*** *Cardiovascular Research and Training Center, University of Alabama in Birmingham, Birmingham, Ala.*

Howard A. Eder, ***Professor of Medicine,*** *Albert Einstein College of Medicine, Bronx, N.Y.*

Judah Folkman, ***Professor of Pediatric Surgery and Professor of Anatomy,*** *Harvard Medical School, and Senior Associate in Surgery, Children's Hospital, Boston, Mass.*

Robert Furchgott, ***Professor of Pharmacology,*** *Downstate Medical Center, Brooklyn, N.Y.*

John C. Hoak, ***Professor and Chairman,*** *Department of Medicine, University of Vermont, Burlington, Vt.*

Eugene Renkin, ***Professor and Chairman,*** *Department of Human Physiology, School of Medicine, University of California, Davis, Calif.*

John Repine, ***Professor of Medicine and Pediatrics,*** *Webb-Waring Lung Institute, Denver, Colo.*

Russell Ross, ***Professor and Chairman,*** *Department of Pathology, School of Medicine, University of Washington, Seattle, Wash.*

Una Ryan, ***Professor of Medicine,*** *University of Miami School of Medicine, Miami, Fla.*

Alan D. Schreiber, ***Professor of Medicine,*** *Hospital of the University of Pennsylvania, Philadelphia, Pa.*

Thomas W. Smith, ***Chief,*** *Cardiovascular Division, Brigham and Women's Hospital,* and ***Professor of Medicine,*** *Harvard Medical School, Boston, Mass.*

Peter Ward, ***Interim Dean,*** *University of Michigan Medical School, Ann Arbor, Mich.*

Albert Winegrad, ***Professor of Medicine,*** *Cox Institute, Philadelphia, Pa.*

Staff

Barbara Filner, ***Director,*** *Division of Health Sciences Policy, Institute of Medicine*

Barbara Mandula, ***Consultant***

Naomi Hudson, ***Administrative Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON

INTERACTIONS BETWEEN BLOOD AND BLOOD VESSELS (INCLUDING THE BIOLOGY OF ATHEROSCLEROSIS)

In atherosclerosis (a form of hardening of the arteries), blood constituents are deposited in the blood vessel wall, causing clotting and inflammation. If the vessel becomes blocked, the consequences can be extremely serious. Medical care and medical research have long been concerned with the complications of atherosclerosis, ranging from heart attacks and strokes to peripheral vascular insufficiency and kidney disease. Complications of atherosclerosis are relatively common, account for an extraordinary toll of chronic impairment and early deaths, and result in staggering economic costs. Despite a 30 percent decline in death rates from heart attacks and strokes in the United States in the past 15 years, complications of atherosclerosis still cause almost half of all deaths in this country. The disease was identified as the leading cause of days lost from work in 1977 and was estimated to have cost almost $40 billion in health expenditures and lost productivity in the United States.

It is now evident that atherogenesis, the disease process that culminates in atherosclerosis, entails a complex interplay between blood and the vessel walls that contain it. In recent years, great progress has been made in understanding this interplay, and this should enhance our ability to treat and prevent the diseases that result from abnormal blood/vessel interactions. The new knowledge has resulted from interdisciplinary efforts drawing on fresh concepts and new techniques of molecular and cell biology, physics, chemistry, and genetics.

Atherogenesis has been recognized as just one example, albeit one of overriding practical importance, of the disease processes that occur when there is a breakdown of normal interactions between blood constituents and the vessel wall. Studies of such interactions also pull together information about normal and abnormal functions of the circulation, the lungs, and the heart. These studies have the potential to provide useful information for a wide variety of human ills.

This paper considers first the characteristics of the blood vessel wall, especially the endothelium, which is in contact with the blood. Second, the paper discusses the properties of some components of the blood that interact with vessel walls. Third, examples are provided of interactions between blood components and the vessel wall that are of broad biomedical significance. Each of these examples is intended to illustrate normal interplays and the consequences of their disordering by disease or injury. Atherogenesis will be discussed in some detail in the context of recent findings. Finally, opportunities that require further investigation are identified.

THE CIRCULATORY SYSTEM

Each component of the circulatory system has its own distinctive character, metabolism, behavior, and disorders. Blood flowing through this system consists of a clear liquid plasma in which distinctive cells are suspended. The large vessels serve as conduits, but the tiny vessels, principally the capillaries, serve as semipermeable barriers for exchange, between blood plasma and tissues, of oxygen, carbon dioxide, nutrients, wastes, and other substances. Capillaries also have functions peculiar to the tissue or organ in which they are situated. In the lungs, for example, the vast capillary network acts as a physical filter to prevent particles, such as clots, from reaching and harming vital organs, particularly the brain and heart.

Large arteries are subject to atherosclerosis and both arteries and veins incur thrombosis, or clotting. Capillaries can be damaged by inflammatory diseases, such as occurs in the kidney in glomerulonephritis.

Although the architecture and basic structures of the circulatory system have been known for a long time, the detailed molecular mechanisms by which the system carries out its functions are only now being unraveled. The use of modern biologic techniques has made it possible to analyze interactions among the various components of the system and to accrue new information about control mechanisms in health and disease.

ENDOTHELIUM

The inner lining of blood vessels is a single layer of cells, the endothelium, which has remarkable properties. It has an apparently simple structure and once was thought to fulfill fairly simple tasks; now it is known to have biologic functions that are many, complex, and sometimes unique to the organs in which the blood vessels are located. In general, the monolayer serves as a selective barrier to the movement of fluids and cells between blood and tissue, as an antithrombotic lining, and as a metabolic organ in its own right.

The lining of vessels normally allows the passage of blood for a lifetime without clotting and without its constituents becoming denatured. No artificial substitute has yet been created that can do that. But an injury upsets the dynamic balance between the vessel wall and blood constituents. A subtle injury, in the form of sustained high blood pressure or high blood cholesterol levels, appears to enhance the permeability of the vascular wall to plasma constituents and enables cells to migrate from the blood stream and possibly contribute to the beginnings of disease.

Information transfer among the components of the vessel wall, and between the endothelial lining and blood constituents, is prerequisite for normal structure and function of the circulatory system. In capillaries, whose wall consists entirely of endothelium, communication between contiguous cells somehow ensures that the vessel lining is a continuous monolayer. In arteries, whose walls are more complex, endothelium communicates with muscle cells that form the next layer outside of it, generally influencing the state of contraction or dilation of the vessel. How this cell-to-cell communication operates is of great investigative interest, and the results of current studies may radically change traditional concepts of how blood flow and pressure are controlled.

Endothelium as a Metabolic Organ

New techniques of culturing endothelial cells have made it possible to study their biochemical activity under controlled conditions and to harvest sufficient material for biochemical study. A combination of in vivo and in vitro studies, for example, has disclosed the vital role of pulmonary capillary endothelium in maintaining systemic blood pressure. On the luminal surface of pulmonary capillary endothelium, the surface in contact with blood, angiotensin-converting enzyme degrades bradykinin (a vasodilator) to inactive products. The same enzyme also converts angiotensin I to angiotensin II (a vasoconstrictor), thereby further helping to sustain systemic blood pressure. In addition, pulmonary capillary endothelium generates angiotensin III, a hormone that acts on the kidneys to influence blood pressure by affecting salt and water balance.

Endothelial cells also participate in maintaining the fluidity of the blood while ensuring that clot formation occurs locally as needed to limit bleeding. Sequences of enzymatic reactions ("cascades") have been identified that either form clots or dissolve clots. Cellular components of the blood, particularly the platelets, also contain powerful substances that lead to clot formation when they encounter injured endothelial lining. The endothelial lining locally generates a hormone (prostacyclin) that modulates the clotting response. Bradykinin, released and destroyed locally, enhances the generation of prostacyclin. These are complicated interplays, regulating the coagulation-anticoagulation systems. Much remains to be learned about the coordinated responses, the mechanisms that prevent under- and overreaction during injury or stress, and the imbalances that arise from kidney failure, or a variety of prescribed drugs, or inherited disorders involving these components.

The growing understanding of endothelial functions has been paralleled by insights into its structural complexities. Ultrastructural specialization of the endothelial cell surface occurs, allowing it to localize diverse functions and carry them out more effectively. Some domains of functional activity are distinctive because of electrostatic charge, while others occur as invaginations in which enzymes are concentrated. How the various molecular activities are organized and orchestrated is currently being explored in depth.

Further study is needed to compare the behavior of endothelial cells in vivo and in vitro, to determine optimal methods for growing endothelial cells from different sites and organs in vitro so as to simulate natural conditions, and to compare endothelial function in large and small vessels. Endothelial cells in culture also are beginning to be used to produce biologically useful substances, such as plasminogen activator, which is important in clot dissolution, and to test pharmacologic agents for toxicity. This research also should continue to be productive.

Transport of Material Across Endothelial Cells

Substances must cross the endothelium in order to leave the blood and be delivered to tissues. Scientists are using a variety of techniques to explore the mechanisms involved in the passage of substances between blood and tissues and the reasons for the different structures of endothelium in different parts of the body.

Endothelium that lines the capillaries of muscle and lungs is continuous; the cells of the monolayer are linked by tight junctions. In contrast, endothelium that lines liver sinusoids and kidney glomerular capillaries has gaps between adjacent endothelial cells so that the adjacent basement membrane, rather than the endothelial layer, constitutes the initial barrier between blood and tissue.

Until recently, the passage of large molecules, such as proteins, was explained in terms of a "pore theory," which envisioned movement through holes or tunnels in the capillary wall. However, experimental observations have made this theory less convincing. At present, there is particular interest in the biochemical attributes of the endothelial cell and the underlying interstitial space as determinants of the transcapillary passage of blood constituents. In addition, attempts continue to relate anatomical structures, such as vesicles, to the transcapillary passage of charged and uncharged particles. The rapidly developing field of transendothelial transport has added fresh impetus and direction to these studies.

Communication Between Endothelium and Smooth Muscle

The state of contraction of the smooth muscle in the walls of small arteries and arterioles largely determines blood pressure and is responsible for adequately distributing blood to all parts of the body. A variety of natural regulatory mechanisms automatically adjust the state of contraction, or vascular tone. Some pharmacologic agents act directly to increase or decrease vascular tone, that is, to cause vasoconstriction or vasodilation.

Only recently has it been demonstrated that certain agents act as vasodilators by acting on the endothelium rather than on the muscle of the vessel wall. It has been proposed that a humoral substance, "endothelium-derived relaxing factor" (EDRF), mediates this dilation. Biochemical substances that seem to exert a vasodilating effect through EDRF include acetylcholine, substance P, adenosine diphosphate, and adenosine triphosphate. The finding has important implications for local regulation of blood supply both under normal conditions and under conditions of vascular disease in which physical and chemical links between cells of the vessel wall are often markedly deranged.

The chemical identity of EDRF has not yet been established. It is possible that EDRF has a central role in maintaining proper blood pressure and that further study of it will help in clarifying and controlling abnormalities in blood pressure.

Angiogenesis

The growth of new capillaries is greatly dependent on the activity of endothelial cells. The endothelial cells that line capillaries are usually in a resting state. Their rate of proliferation is so low that turnover times are measured in years. However, these endothelial cells are capable of sudden growth spurts during periods when new capillaries are rapidly forming, for example, during wound healing. The process of new capillary formation is called "angiogenesis," and cell turnover times during angiogenesis are measured in days.

Protracted angiogenesis is observed in such diverse pathologic states as arthritis, psoriasis, diabetic retinopathy, and chronic inflammation; the most sustained and intense angiogenesis is associated with solid tumors. In many of these abnormalities, angiogenesis itself contributes to the disease process. In arthritis, new capillaries can invade and destroy joint cartilage. In diabetes, new capillaries in the retina of the eye hemorrhage and cause blindness. Progressive growth of tumors and their ability to form metastases depend upon continuous angiogenesis.

Because of techniques introduced since 1970, it is now possible to investigate angiogenesis experimentally. Capillary endothelial cells have been cloned and grown in vitro, where they form tubes, branches, and whole capillary networks. Special polymers have been developed for the sustained release of angiogenic factors in vivo. Bioassays for angiogenic activity have been perfected in the chick embryo and in the rabbit and mouse cornea.

Application of these techniques is revealing some components of the angiogenic process. As an early step in capillary formation, endothelial cells release specific enzymes that punch holes in the vascular basement membrane. Endothelial cells migrate through these holes. The cells align in a linear configuration to form a new sprout. The sprout elongates through interstitial tissue toward an angiogenic stimulus such as a small tumor. Sprout elongation depends upon chemotaxis of endothelial cells at the tip of the sprout, division of cells farther back, and production of interstitial collagenases, probably by all of the endothelial cells in the sprout.

Techniques developed primarily for the study of tumor angiogenesis have recently been used to explore the mechanisms of angiogenesis associated with processes such as wound healing. Endothelial cells and debris-scavenging macrophages play major roles. In the depth of a fresh wound, tissue oxygen levels are extremely low because capillaries have been damaged and blood flow interrupted. In a setting that is poor in oxygen, macrophages that migrate into the wound secrete angiogenesis factors. As new capillaries are attracted into the wound, oxygen tension rises; the macrophages then reduce or shut off their stimulation of angiogenesis. The details of this feedback mechanism remain to be worked out, and the macrophage angiogenic factor has not yet been purified.

BLOOD CONSTITUENTS

Blood is composed of circulating cells suspended in a liquid, the plasma. Each component of blood has access to endothelium and virtually all bear electrostatic charges that interact with each other and with the surface charges on the endothelial lining. The composition of the blood is automatically adjusted to the various needs of the body, including respiratory gas exchange, maintenance of blood pressure, and maintenance of fluid and electrolyte balance. Blood constituents interact with each other as well as with the vessel wall. A few of the blood constituents that have crucial roles in interacting with the vessel wall are discussed below.

Platelets

Platelets are essential elements in blood clotting and also have a central role in the process of atherogenesis. Upon contact with collagen or macrophages in an area of vascular injury, platelets aggregate and their storage granules discharge biologically active substances that promote the formation and organization of a clot to plug the breach in the endothelial lining. Enzymes and growth factors released by the aggregated platelets attract other platelets, evoke spasm and proliferation of adjacent vascular smooth muscle, and draw coagulation proteins to the area for the assembly of the clot. These reactions are controlled by molecular feedback mechanisms. For instance, the activity of thromboxane, released by platelets to promote local vascular constriction and accumulation of other platelets, is modulated in its clot-promoting effects by prostacyclin (an anti-aggregating and vasodilator agent) liberated by endothelium.

New techniques have greatly expanded our knowledge of platelet structure and function. The surface of the platelet displays specialized molecular organization, which includes receptors that provide binding sites for macromolecules that engage in complex immunologic reactions, and other receptors that have to do with the clotting process and that bind fibrinogen and von Willebrand factor. These receptors constitute ultrastructural foci for the orderly assembly of the clot. As the coagulation proteins fall into place on the platelet surface, the clotting process is dramatically accelerated by a surge in formation of thrombin.

There are still large gaps in the understanding of how platelets engage in their complex interactions with the vessel wall, plasma factors, and other blood cells. Their role in the inflammatory process, for example, remains enigmatic. Platelets, in addition, are a storehouse for powerful enzymes that differ strikingly in structure and function. How they interrelate in response to injury requires much further study.

Leukocytes

Leukocytes, or white blood cells, are part of the body's defense system. They are drawn chemotactically to areas of damage or invasion by foreign agents and participate in the ingestion, walling off, and elimination of the offending agent. Upon stimulation by appropriate chemicals, leukocytes generate oxygen radicals and release substances capable of destroying microorganisms. We are now beginning to understand how the body protects itself against inappropriate activity of these powerful chemicals. Phagocytic cells involved in the inflammatory response, as well as other cells of the body, are equipped with intracellular antioxidant enzymes, such as catalase and glutathione peroxidase; in addition, there are circulating neutralizers, such as ceruloplasmin and alpha-1 antitrypsin, to prevent these chemicals from damaging the host when they are released. It is anticipated that awareness of how leukocytes inflict their damage will make it possible to develop medications to protect people against their toxic products.

Plasma Lipoproteins

Because lipids are insoluble in an aqueous medium, they are transported in plasma as macromolecular complexes called lipoproteins. Of the various types of lipoproteins, low-density lipoprotein (LDL) carries one-half to two-thirds of all the cholesterol in plasma. LDL has cholesterol ester as its core lipid and unesterified cholesterol on its surface. High levels of LDL-cholesterol predispose to atherosclerosis. But another form of lipoprotein, high-density lipoprotein (HDL), may be anti-atherogenic.

Cells of many different tissues have specific receptors for LDL and transport it into the cytoplasm, where it subsequently is degraded in lysosomes. In normal circumstances, however, the entry of LDL into the artery wall is predominantly receptor-independent and occurs by a nonspecific process known as transcytosis. Reverse transport of LDL also occurs and avoids undue accumulation of cholesterol within the cell. The mechanisms involved in reverse transport and in the feedback control of intracellular cholesterol concentrations are under investigation, as is the mechanism of LDL transport into and out of endothelial cells following injury to the vascular lining.

Macrophages and endothelial cells have specific receptors for LDL, although their rate of uptake of LDL is low. Macrophages also have specific receptors for LDL that has been chemically modified or altered by contact with endothelial cells. The rate of uptake of modified LDL by this receptor is considerably higher than that of unmodified LDL, and foam cells (macrophages filled with cholesterol) can be readily produced by interaction of modified LDL with macrophages. LDL also apparently alters endothelial cell membranes in a way that allows access of lipoproteins into the vessel wall. When LDL is present in the vessel wall, it may cause proliferation of smooth muscle cells, which migrate into the lining of the blood vessel of the affected artery.

Considerably less is known about the interaction of vessels with the larger lipid-rich lipoproteins, such as chylomicrons. These are triglyceride-rich particles that appear in the circulation after the ingestion of fat and in certain disease states. These are removed by the liver, which then secretes the cholesterol in the particles in the form of VLDL, which is metabolized in part to LDL. It is by this mechanism that dietary cholesterol eventually enters the vessel wall.

In persons with a disorder called Type III hyperlipoproteinemia, atherosclerosis occurs frequently. In these patients, the chylomicron remnants and VLDL remnants enriched in cholesterol accumulate in the plasma in a form called beta-VLDL. Macrophages from patients with Type III hyperlipoproteinemia have receptors for beta-VLDL, which probably accounts for the accumulation of foam cells in vessel walls in the distinctive skin lesions of these patients.

Earlier studies had suggested that high-density lipoproteins (HDL) are involved in transporting cholesterol out of macrophages. HDL contains a number of apoproteins (the protein portion) of lipoproteins. The major one in HDL is apo A-I. A specific receptor for apo A-I has been found on endothelial cells, on vascular smooth muscle cells, and on some liver cells. It has been demonstrated recently that the apo A-I receptor on nonliver cells is regulated by the cholesterol content of the cells. Thus, HDL may transport intracellular cholesterol out of the vessel, becoming enriched in cholesterol while doing so. How the receptor for HDL on liver cells is regulated is not yet known.

Mechanisms of entry of LDL into the vessel wall and handling of LDL when endothelial cells are damaged require further investigation. The extent of interaction of chylomicron remnants or VLDL remnants with the vessel wall also needs study. An exciting new area of investigation involves the mechanism by which HDL, which can bind to specific receptors on vascular endothelial cells and on smooth muscle cells, transports cholesterol from these tissues. New techniques, such as production of monoclonal antibodies for the various apolipoproteins, will be important for studying the metabolism of apolipoproteins and their role in atherogenesis.

EXAMPLES OF BIOMEDICALLY IMPORTANT INTERACTIONS

Molecular interactions between the constituents of blood and the endothelium operate continuously to permit normal growth and development and to sustain the metabolic activities of daily life. Respiratory gases, water, nutrients, and other molecules must move between endothelium, plasma, and the cells of the blood. Electrolytes pass back and forth between blood and tissues where they trigger and drive metabolic machinery. Hormones travel long distances in blood to activate and participate in biological processes far removed from their sites of origin. Each of these systems is governed by elaborate control mechanisms that continuously monitor and restore dynamic equilibria. To illustrate the operation, complexity, and control of the biologic systems and their regulatory mechanisms, three examples will be considered: blood clotting, the inflammatory process, and atherogenesis.

BLOOD CLOTTING (THROMBOSIS)

The tendency for blood to clot when a blood vessel is severed is a safeguard against bleeding to death. In living systems, this tendency is counterbalanced by mechanisms to ensure blood fluidity; inappropriate or excess clotting may lead to serious clinical syndromes. Thrombosis plays a critical role in initiating, propagating, and occluding the atherosclerotic lesion, thereby constituting the precipitating cause of acute myocardial infarction, stroke, and peripheral vascular disease. Thrombosis is a precursor of thromboembolic disease, a common and life-threatening complication of surgery, heart failure, and cardiac arrhythmias. Improved understanding of thrombotic events—involving endothelium and blood constituents—holds the prospect of better therapeutic and prophylactic interventions.

The complex sets of reactions involved in clot formation and removal have been examined experimentally for years, but more pieces of the puzzle continue to be found. For example, one recent discovery is Protein C, a vitamin-K-dependent coagulation factor and a potent anticoagulant and fibrinolytic agent. Its concentration is regulated by an elaborate biochemical control system: Protein C is activated by thrombin bound to the protein thrombomodulin on endothelium; activated Protein C is neutralized by an inhibitor present in the plasma. Protein C appears to be crucial for life and several infants who lacked Protein C because of a genetic defect died of generalized thrombosis shortly after birth.

THE INFLAMMATORY PROCESS

The inflammatory process is a vital component of the defense mechanisms of the body. A cellular response, abetted by an increase in endothelial permeability evoked by inflammatory peptides and lipids, contributes to the leakage of water, salts, and proteins from plasma into the extravascular environment. Inherent in this process is the migration of leukocytes from the blood to the site of injury. En route, these blood cells traverse endothelium.

New concepts and techniques are currently being applied to the interplay between the blood and endothelium in the inflammatory process. Promising lines of research include the effects of particular types of endothelial injury on the inflammatory response, the response of endothelium to substances, such as bradykinin, that are involved in the inflammatory response, and the mechanisms that promote adherence between neutrophils (a type of white blood cell) and the endothelial cell.

ATHEROGENESIS

Atherosclerosis is a special type of arteriosclerosis that affects large arteries, such as those of the coronary and cerebral circulations. Atherogenesis is the process that culminates in atherosclerosis. It results from misdirections and overresponses in normal biologic interplay at the molecular level. In its early stages—those most relevant to the understanding of initiating mechanisms and the institution of preventive measures—the inner layer of an affected artery is thickened by a deposition of fat, particularly cholesterol, and by an increase in the number of cells in the lining of the vessel. Two hypotheses concerning the etiology and pathogenesis of atherosclerosis continue to be studied intensively: the lipid infiltration hypothesis and the response-to-injury hypothesis. Each of these has stimulated new knowledge about the interplay between the vascular wall and the blood in atherogenesis.

Lipoproteins in Atherosclerosis

As noted above, the importance of lipoproteins as the cholesterol-carrying substances in blood has been clarified during the last decade. The recent decrease in mortality from heart attacks and strokes has been attributed, in part, to improved understanding of the lipoproteins and their relation to risk factors, such as high blood pressure, smoking, diabetes, and high levels of blood cholesterol. However, still to be resolved is how risk factors act as injurious agents and in what ways the responses to these diverse types of injury differ. New approaches to slowing, arresting, and reversing the atherosclerotic process are being sought. Among the more promising is modification of diet, coupled as needed with the use of medications designed to lower blood levels of cholesterol. One recent lead is the possible effect of certain fats present in seafood in lowering blood levels of cholesterol and low-density lipoproteins.

Response-to-Injury Hypothesis

This hypothesis for atherogenesis, proposed in 1973, suggested that the various risk factors associated with atherosclerosis somehow lead to alterations in the endothelial cells, and that these represent various forms of "injury."

An important development associated with the response-to-injury hypothesis has been the discovery of growth factors that may be important in the development of the lesions of atherosclerosis. Two of these are the platelet-derived growth factor (PDGF) and the macrophage-derived growth factor. PDGF is the principal substance in whole blood serum that is responsible for proliferation of cells, such as smooth muscle and fibroblasts, in cell culture. PDGF may be associated with initiation and progression of the lesions of atherosclerosis, because experimentally induced lesions can be prevented by agents that substantially decrease circulating platelets or inhibit platelet interactions.

PDGF plays an important role in both mobilization and proliferation of cells. It has been purified to homogeneity, characterized, and sequenced. PDGF binds to a specific high-affinity receptor on the surface of susceptible cells, and promptly induces such intracellular events as tyrosine phosphorylation of several proteins, including the receptor for PDGF; phospholipase activation leading to diglyceride formation; and formation of arachidonic acid and prostaglandin. It also leads to increased binding of lipoproteins, increased endocytosis, increased protein and RNA synthesis, and then, after 24 to 36 hours, to DNA replication and cell division. Platelet-derived growth factor also is a chemotactic agent that can attract cells, such as smooth muscle, into the cavity of the affected artery.

Experiments concerning the cellular and molecular biology of endothelium (which is also a potential source of platelet-derived growth factor), smooth muscle, the monocyte/macrophage, and the platelet will be important in understanding the lesions of atherosclerosis. Studies that examine the interactions of these cells have opened new frontiers and represent vitally important areas for further research.

Thrombosis in Atherosclerosis

Advances in our knowledge of blood clotting factors, platelets, and prostaglandins have stimulated new insights into old theories about how thrombosis is involved in all stages of atherosclerosis. Studies have demonstrated that platelets, the principal cells involved in thrombosis, may play a role in the initiation of atherosclerosis. Such studies may help to control and prevent thrombosis.

IMPLICATIONS AND PROSPECTS

The concepts and techniques of cell biology that deal with structure, function, and control mechanisms at the molecular level are now available for application to the important biomedical problem of interactions between the vessel and the blood that flows through it. To date, the major focus has been on the individual components involved in this interplay. The time is now ripe for a concerted interdisciplinary effort to understand the control mechanisms that regulate the interplay in health and the derangements arising from perturbations in the system arising from injury and disease.

Favoring a concerted effort at this time is the current capability for harvesting large quantities of endothelium grown in culture. In addition, innovative cell sorters and isolation techniques make it possible to isolate manageable quantities of individual blood cells for manipulation and study. Biochemical and biophysical techniques also provide precise tools for quantifying constituents in plasma. Finally, receptors and channels at the level of the cell and its organelles can be examined using approaches of molecular and cell biology that were previously unavailable.

The practical rewards of this fundamental research have been suggested in the sections of this paper devoted to thrombosis, inflammation, and atherosclerosis. Many other prospects can be reasonably expected to follow greater understanding of the blood and blood vessel interactions: improved prosthetic devices, such as heart valves, vascular replacements, dialyzing membranes, and artificial organs; fresh insights into the prevention and treatment of vascular occlusions and their complications; new approaches to the prevention, diagnosis, management, and reversal of atherosclerotic lesions. Study of interactions between the blood and the blood vessels provides a dramatic opportunity for progress in fundamental science and holds extraordinary promise for practical application.

RESEARCH BRIEFING PANEL ON THE BIOLOGY OF PARASITISM

John R. David ***(Co-chairman), John LaPorte Given Professor and Chairman,*** *Department of Tropical Public Health, Harvard School of Public Health, and Professor of Medicine, Harvard Medical School, Boston, Mass.*

Joshua Lederberg ***(Co-chairman), President,*** *The Rockefeller University, New York, N.Y.*

George Cross, ***Professor of Molecular Parasitology,*** *The Rockefeller University, New York, N.Y.*

Paul Englund, ***Professor of Biological Chemistry,*** *The Johns Hopkins University School of Medicine, Baltimore, Md.*

Paul Gross, ***President and Director,*** *Marine Biological Laboratory, Woods Hole, Mass.*

Seymour J. Klebanoff, ***Professor of Medicine,*** *and* ***Head,*** *Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington, Seattle, Wash.*

Adel Mahmoud, ***Professor of Medicine, Chief,*** *Division of Geographic Medicine, Case Western Reserve University, University Hospital, Cleveland, Ohio*

Adolfo Martinez-Palomo, ***Head,*** *Section of Experimental Pathology, Centro de Investigacion y de Estudios Avanzados del Instituto Politechnico National, Mexico City, Mexico*

Staff

Barbara Filner, ***Director,*** *Division of Health Sciences Policy, Institute of Medicine*

Barbara Mandula, ***Consultant***

Naomi Hudson, ***Administrative Secretary***

REPORT OF THE RESEARCH BRIEFING PANEL ON THE BIOLOGY OF PARASITISM

Parasites are organisms that require another organism, a "host," for survival and that usually do some harm to the host. The field of parasitology has been limited to those organisms that belong to the animal kingdom, although bacteria, viruses, and fungi are also parasites. Of concern here are the single-celled protozoans, such as cause malaria and leishmania; multicellular worms or helminths, such as schistosomes and filaria; and arthropods, which include insects. Many of the arthropods also are vectors that transmit the protozoan and helminthic parasites.

There are two major reasons for the current intense interest in the study of parasites. First, such studies will have a large impact on our understanding of basic biologic processes such as those involved in cell growth and differentiation. Many of the human parasites have multiple hosts and may have several developmental stages in each host. Each stage involves regulated transitions from one form to another. Further, parasites have evolved very specialized adaptations that enable them to infect the host and then evade the host's defense mechanisms while they survive and replicate, utilizing for these purposes the host's metabolic processes. The mechanisms by which parasites adapt to their environment are diverse and ingenious, and are only beginning to be understood.

Second, application of the knowledge obtained in these basic investigations promises to have a major impact on world health. Of the six diseases singled out for research emphasis by the Special Program for Research and Training in Tropical Diseases, sponsored by the World Bank, the United Nations Development Program, and the World Health Organization, five are caused by parasites: malaria, schistosomiasis, filariasis, trypanosomiasis, and leishmaniasis. These and other parasites affect more than a billion people worldwide. Approximately 300 million suffer from malaria, 200 million have schistosomiasis, and 300 million have filariasis. Another 40 million have onchocerciasis, which causes "river blindness."

Parasite diseases also affect the United States. Giardia, a protozoan parasite, is one of the most common causes of epidemic infectious diarrhea in this country. In addition, we must consider the immigration of a large number of people with parasitic diseases, the vast outflow of American tourists and business representatives overseas, and the potential exposure of military personnel in countries where parasitic infections are common.

The potential rewards of studying parasites are enormous, both in terms of learning about basic biologic processes and in combating disease. Because of their complexity, animal parasites are difficult to study. However, modern biologic concepts and advances in techniques are now enabling researchers to address such exciting and long-standing questions about parasite-host interactions as the following: (1) How are these unicellular and multicellular organisms capable of establishing infections? (2) What genes regulate the transformation of parasites through their complex life cycle? (3) What are the mechanisms of disease syndromes that result from parasitic infections? (4) What is the array of possible host protective responses and can these be enhanced? (5) What mechanisms do parasites use to evade the host's immune and other protective defenses? (6) Are vaccines feasible? (7) What biochemical pathways do parasites have that differ from the host and can these be used as targets for new drugs? (8) What is the basis for the ability of certain insects to transmit parasites or to become resistant to insecticides, and can these properties be altered?

Some of the technical advances that have been crucial to the study of parasites are the production of monoclonal antibodies, the isolation of specific genes, and the culturing and growth of some of these organisms in the laboratory. This paper describes some basic biologic questions that are being studied in parasites by several different scientific disciplines, and then suggests potential future applications of such information to prevention and treatment of parasitic diseases.

BASIC RESEARCH ON PARASITES

MOLECULAR BIOLOGY

Antigenic Variation

African trypanosomiasis is characterized by cyclic parasitemia, that is, by the appearance in the blood stream every 7 to 10 days of waves of organisms. The surface of the protozoan's membrane is covered by a single type of glycoprotein, which is antigenic, and the parasites of each cycle contain a completely different glycoprotein. Although the body mounts an antibody response, the parasite can change its surface antigen hundreds of times, always keeping ahead of the host's immune response.

What is the basis for the capacity of these parasites to alter their surface antigens? Investigation of this phenomenon, undoubtedly the most sophisticated mechanism yet devised for evading the immune response of the host, uncovered the first example of "jumping genes" for surface proteins. Each organism has an estimated 300 to 1,000 different genes coding for these variant surface glycoproteins, but only one is expressed at a time. The gene to be expressed is duplicated and then moved to another part of the chromosome where it is expressed. It is not yet known how these parasite genes are selected, how they are turned on and off, or how they are regulated. Nevertheless, studies of the molecular biology of the African trypanosome already have had considerable influence on our understanding of gene rearrangement in general, and promise to uncover mechanisms of gene regulation.

Surface Antigen Repeating Units

Sporozoites, the form of the malaria parasite that matures in the salivary gland of the anopheles mosquito and is injected into the human skin during a bite, are free in the human host's blood for only a few minutes before entering liver cells. Because of their precise specificity for a single antigen, monoclonal antibodies have allowed the detection and characterization of surface proteins on sporozoites. The antibodies are specific to a given species as well as to that stage in the life cycle.

The antibodies directed against these surface antigens neutralize sporozoite infectivity, so there has been great interest in both the antibodies and the antigens they have identified.

Genes coding for the sporozoite surface antigens have been introduced into bacteria by recombinant DNA techniques. This allows the production of sufficient antigen to study it in detail. The antigens have been found to have unusual repeating sequences. In the sporozoite of the malaria species *Plasmodium knowlesi*, for example, the unit is made up of a peptide with 12 amino acids repeated 12 times in tandem. The antigen of *Plasmodium falciparum* has a different repeating unit, with 4 amino acids repeated 23 times. It will be interesting to learn how common such repeating antigen units are in other parasites and what evolutionary pressures may have led to their development. In a more practical vein, such small reactive units lend themselves to the production of synthetic peptide antigens that may be used to develop vaccines effective against malaria.

Kinetoplast DNA

Kinetoplast DNA (kDNA) is a form of mitochondrial DNA found in some protozoans such as trypanosomes and leishmania, which is essential for their survival. It has a remarkable structure consisting of thousands of DNA circles interlocked in a network. A minor portion of these circles, the maxicircles, code for mitochondrial proteins. The majority, the minicircles, are only about 1,000 base pairs in size. These do not appear to code for any protein, and their function is not known. Minicircles are the only DNA in nature known to have a region of bent DNA helix. It is of great interest to learn both the function of this unusual DNA conformation and how its nucleotide sequence induces the curvature.

Gene Regulation During Life Cycle Transformations

Parasites undergo profound changes during the various stages of their life cycles, and work on the molecular biologic basis of these changes is in its infancy. The next few years should see dramatic progress in work on the molecular basis for developmental transformation in many different protozoan and helminthic parasites.

Leishmania, for example, go from a flagellated protozoan form in the insect host to a smaller form without flagella when in a mammalian macrophage. This change is reversed after the protozoa are taken up by a fly when it bites an infected animal. The amount of tubulin, a structural protein of the cytoskeleton, decreases as the parasite goes from the flagellated to the nonflagellated form and then increases when the parasite returns to the flagellated form. In this case of transformation, the control of the genes for tubulin occurs at the level of messenger RNA processing.

IMMUNOLOGY

Studies in immunology have concentrated in several areas. These include mechanisms the host can muster that the parasite cannot escape, mechanisms of immune evasion, and the role of the immune response in causing tissue damage. The information obtained has been useful in developing better diagnostic reagents and forms the basis for the development of protective vaccines.

Protective Mechanisms of the Host

Studies on host responses to parasites have revealed some novel systems. For example, one such system was found while studying the antibody-dependent ability of host cells to kill schistosomes. These studies led to the first demonstration that the eosinophil, a white blood cell, can act as a killer cell. These findings have now been applied to other areas where it has been shown that eosinophils play an important part in producing tissue damage in certain types of heart disease, in inflammation, and in allergic diseases such as in asthma.

As another example, work on the immune response to ticks has highlighted the role of basophils, another type of white blood cell, and mast cells in producing vascular permeability in the early stages of cell-mediated immune reactions. Furthermore, studies of the schistosome-linked granuloma (an inflammatory nodule) were the first to demonstrate suppression of lymphocyte hormone production by T-cells. (T-cells and lymphocyte hormones are important regulatory components of the immune system.)

Mechanisms of Immune Evasion

Parasites have developed many ways of evading the human host's defenses. Some change their surface antigens, some masquerade as the host by taking up host molecules onto their surfaces, and others simply shed their surface antigens. Some have enzymes on their surface that destroy antibodies. Macrophages are the body's major cell for engulfing foreign particles, and they usually kill the organisms they engulf. Some parasites have devised ways of avoiding the effects of the toxic substances present in the lysosomes of the macrophage, although how they do this is not known. Yet other parasites suppress the host's immune attack.

Although at present most of our knowledge is simply descriptive, future studies should elucidate the mechanisms for accomplishing all these forms of evasion. Information about the effects of parasites on the immune system would undoubtedly enhance our understanding of how the immune system functions in many other types of diseases.

Immunopathology

Studies on the mechanisms of tissue damage in schistosomiasis demonstrated that the host reactions to egg antigens are the main factors producing pathology. Further studies on this process were the first to show that mononuclear cells of the host release factors that promote fibrosis. The mechanisms of pathology induced by many different parasites merit further study. Of particular interest are the ways parasites induce autoimmune reactions such as those postulated in infections caused by *Trypanosoma cruzi,* the mechanism of blindness caused by *Onchocerca volvulus,* the process of sequestration of infected red blood cells leading to malaria damage in the brain, and the underlying basis for destruction of cartilage in mucocutaneous leishmaniasis.

MEMBRANE BIOLOGY AND CELL BIOLOGY

Unusual Features of Parasitic Membranes

The first interactions between the parasite and the host are membrane-membrane interactions. Many parasites have membranes that are different from those in mammalian cells, and considerable effort has been directed at learning about the function and structure of parasite membranes. As a result, several interesting discoveries have emerged.

One example is the finding that the variant surface glycoprotein of African trypanosomes is attached to the membrane by a novel protein-lipid bond. A newly recognized enzyme is involved in destroying this bond. Another unusual finding concerns an enzyme on the surface membrane of *Trypanosoma cruzi.* This organism has a neuraminidase, an enzyme that cleaves off the most important sugar group on the surface of mammalian cells, namely, sialic acid. The role of this enzyme should be studied, especially as it relates to parasite survival and tissue injury. The question of whether similar enzymes are present on other parasites also is of interest.

Parasite-Host Membrane Interactions

Work is being carried out on the mechanism by which malarial merozoites (the mobile infective stage) invade red blood cells. Glycophorin, a glycoprotein on the red blood cell surface, appears to be an important part of the receptor for the parasite. How parasites disrupt the rigid cytoskeleton of the red blood cell so they can enter it is unknown, as is the source for the special membrane that surrounds the parasite once it is in the red blood cell. Furthermore, parasite antigens soon appear on the outer surface of the red blood cell. To reach the surface, parasite antigens must traverse the parasite's own membrane, then the special membrane surrounding the parasite, and finally the red blood cell membrane. The mechanism for accomplishing this voyage is unknown.

Alteration of the red blood cell surface is important in enabling the red blood cells that contain the parasite of *Plasmodium falciparum*, the most common of the four malaria species, to lodge in the vascular bed of certain organs such as the brain. These surface antigens also undergo antigenic variation by unknown mechanisms. The spleen can alter the expression of these parasite antigens on the red blood cell surface, but how it does so is another mystery to be solved.

Of interest is the recent discovery that *Entamoeba histolytica*, another protozoan parasite, kills host cells by injecting a protein into the host cell membrane that causes an ion flux and subsequent osmotic lysis of the target cell. These findings have stimulated a search to see whether analogous proteins are involved in cytotoxicity induced by T-lymphocytes.

Clearly, studies on these membrane effects will have ramifications affecting much more than our knowledge of the parasites themselves.

BIOLOGY OF INSECT VECTORS

An attack on the insects that transmit particular pathogens remains a major focus for most programs aiming to control parasitic diseases. Although chemical insecticides provide the major tool for this purpose, their effectiveness has steadily diminished because of insect resistance.

New technologies should lead to revised strategies for combating insect vectors. Recent developments in molecular biology, combined with improved understanding of insect physiology and behavior, should enable us to better understand the mechanisms of insecticide resistance, vector competence—the ability of an insect to transmit a parasite

—and the control, feeding behavior, and reproduction of insects. Ultimately, this technology may enable us to genetically modify insect populations in nature, so that the targeted population is less capable of transmitting pathogens or less susceptible to becoming resistant to insecticides.

Such research, both basic and applied, should have far-reaching benefits, affecting not only the transmission of known disease but also important aspects of agriculture.

BIOCHEMISTRY AND PHARMACOLOGY

The very nature of parasitism implies that the parasite has metabolic needs that the host supplies and, therefore, that the parasite must have metabolic pathways that differ from those of the host. Novel metabolic pathways and even new cell organelles are being discovered. These have led to the design of compounds that can inhibit these unique pathways and thus control some parasites and insects without affecting their hosts. A few examples are described below.

Unusual Metabolic Pathways

In general, parasitic protozoans cannot synthesize purines, which are chemical precursors of nucleic acids. They have therefore developed elaborate pathways to utilize purines from the host. Many of these pathways differ from one protozoan to another. Study of these pathways has greatly enhanced our knowledge of purine metabolism in general and may lead to development of drugs that can kill protozoans but are not toxic to mammalian cells. Polyamines also are required for DNA replication and cell differentiation. Trypanosomes and leishmania have simplified metabolic pathways for biosynthesis of polyamines, which may be susceptible to specific inhibition.

Unusual Organelles

Parasites have been found to contain several novel organelles. One of these, the glycosome, found in trypanosomes, contains the enzymes for glycolysis; its discovery encourages research toward development of specific inhibitors. Another novel organelle is the hydrogenosome found in trichomonas. This organelle contains pyruvate kinase and pyruvate dehydrogenase, two enzymes used in producing the ATP needed by the cell. These hydrogenosomes chemically reduce the drug metronidazol to a toxic compound that kills the parasite. The host cells, which lack this organelle, are spared.

The Neuromuscular Junction

The neuromuscular junction of certain helminths contains receptors for GABA (gamma-aminobutyric acid), which also is a neurotransmitter in the human brain. Ivermectin, a drug that enhances the activity of these receptors, has a profound effect on helminths. It causes paralysis that prevents them from moving or eating. It is exceedingly potent, with less than milligrams per acre of pastureland protecting cattle against helminth parasites. Because this drug cannot traverse the blood-brain barrier, it is nontoxic to humans. Further studies using ivermectin have shown that it also acts on the GABA receptor complex of arthropods. Additional studies with this drug should greatly increase our understanding of insect physiology and yield knowledge of GABA receptors in higher organisms.

Juvenile Hormone

Juvenile hormone (ecdysin) is involved in the differentiation and molting of insects and has recently been found in helminths. Further study of the regulation of molting and differentiation of helminths and the role of ecdysin in this process should lead to entirely new approaches to control of some of these parasites.

APPLICATION TO DISEASES

Increased understanding of the basic biology of parasitism should provide many opportunities for combating diseases caused by these organisms. This is especially important now that traditional public health measures that had helped to control some of these diseases are no longer sufficient by themselves. An example of the need for new methods can be seen in the case of malaria. After World War II, there was great hope that this disease could be eradicated by using DDT to kill the mosquitoes in houses and chloroquin to treat infections in people. Indeed, malaria was eradicated from the United States and around the Mediterranean and decreased markedly in Asia and South America. In the early 1960s in Sri Lanka, the number of cases per year fell from more than 1 million to 18. However, within 5 years, there were again almost a million cases, due to a combination of DDT and chloroquin resistance and to the difficulty in maintaining this costly control program. Now malaria has reached serious epidemic proportions in parts of Asia and South America and continues unabated in Africa where it was never controlled due to the behavior of the local mosquitoes, which, instead of landing on the sprayed walls of the house after biting, go outside. Malaria resistant to the present drugs has become an alarming problem.

The belief that parasitic diseases disappear with modernization and industrialization is not always correct. For instance, the prevalence of schistosomiasis has increased hand in hand with the development of hydroelectric dams required for energy and irrigation projects necessary for improved agriculture. The large lakes behind these dams have added thousands of miles of waterfront and have increased the contact between people and the infected snails that transmit this parasite. Another example is the increase in leishmaniasis in the Amazon where people who are expanding towns and building roads come into contact with sandflies that live in the forest and transmit this infection. Applying insecticides to the forests is neither effective nor feasible.

FUTURE RESEARCH OPPORTUNITIES

There is a great need now to broaden the entire base of biologic research on parasites. This is an opportune time because many of the new concepts and biotechnologies make it possible to solve some of the important questions concerning the biology of parasitism. A multidisciplinary approach is required.

Most of the productive research thus far employing the new biotechnologies has been on two parasites, the malaria parasite *P. falciparum* and the African trypanosome of cattle, *T. brucei.* Much less work in molecular biology, for instance, is being done on the African trypanosomes that infect humans, or on other protozoans and the many helminths.

The main reason for the early focus on *P. falciparum* and *T. brucei* is the relative ease, compared with other parasites, with which these two parasites can be cultured and manipulated in the laboratory. There is much excitement over prospects of a vaccine against the infective form of *P. falciparum,* the sporozoite, but little work has been carried out on the other three species of malaria that infect humans. There is also work in progress on the development of vaccines to the other stages of the *P. falciparum* parasite because it is still not clear how effective the vaccine against the sporozoite will be. The sporozoite vaccine has worked in laboratory tests on rodents, but must now be developed for humans. One problem is that the vaccine will have to work completely in the short time that the sporozoite is present in the blood. An effective adjuvant must be identified, one that will allow for a strong enough response for the vaccine to be clinically effective.

It is now important to take advantage of new technologies to study other parasites. This will require establishing cultures of these organisms in vitro, developing suitable animal models, and establishing appropriate life cycles in the laboratory. Such work is time-consuming, may involve many false starts, and requires cooperation of persons in many disciplines and long-term support.

Molecular biology is providing several new methods for generating antigens needed to form the basis of vaccines. These methods include laboratory synthesis of peptides, producing proteins in microorganisms by recombinant DNA techniques, and incorporation of genes coding for protective parasite antigens into the DNA of attenuated viruses used for unrelated vaccines. The essence of the problem is no longer how to produce antigens by recombinant techniques, but which antigens to produce and how to present them effectively to the immune system in clinically acceptable formulations.

Molecular biologic techniques also are being used for better diagnosis. For instance, the nucleotide sequence of the kinetoplast DNA found in certain protozoans varies among the different species. Using DNA hybridization to identify the kDNA, it is now possible to determine in 24 hours instead of 3 months whether a leishmania infection involves a benign or virulent species. Similar techniques should be applicable to other parasites. Furthermore, the production of species-specific monoclonal antibodies against parasites causing malaria, leishmaniasis, schistosomiasis, and other diseases is leading to the development of greatly improved diagnostic reagents and identification of relevant antigens for the development of vaccines.

The study of the genes that regulate the transformation of the parasite through its different stages should result in novel ways of interrupting the life cycle. Studies of the molecular biology of parasites have already broadened our understanding of molecular biology in general. For instance, it was through the study of malaria that an enzyme, mung bean nuclease, was found recently to have the unique property of cutting genes out of DNA without cutting in the coding region, so that genes can be studied and cloned much more easily than before.

Basic immunologic studies of parasites should be expanded to cover several other areas beyond development of vaccines and diagnostic reagents. For instance, the study of the mechanisms of immune evasion should lead to new methods of overcoming the parasites, and those methods should also be adaptable to such other types of organisms as bacteria, viruses, and fungi. Work on immunopathology, especially possible autoimmune mechanisms, is required to assure that the antigens proposed for vaccines are not also those that can induce pathology detrimental to the host. This type of study is important, for instance, in *T. cruzi* infections, because it has been postulated that the heart disease resulting from this infection may be caused by an autoimmune process.

Basic research on the novel biochemical mechanisms that parasites have evolved is currently one of the most neglected aspects of parasitology. Work in this area should lead to the development of new drugs that are needed to treat some of these diseases. For instance, the major drug for treating African trypanosomiasis is Suramin, first synthesized in 1917. It is a very toxic drug and is effective only in the early stages of the disease, before the parasite reaches the brain and produces the symptoms that give the disease its common name, sleeping sickness. There are no drugs to treat *T. cruzi* infection, and those available against the leishmania organisms were first produced decades ago and are quite toxic.

Studies of the biology of insects that carry and transmit parasites should be expanded to include the application of molecular biology. Alternative methods of insect control, including the genetic modification of insect populations in nature, should result. Investigations in this area should not only help control some of these parasite diseases but also should be applicable to problems in agriculture and animal husbandry.

Despite the enormous potential of research on the basic biology of parasites, relatively few scientists are working in this area, and many more should be encouraged to do so. The foregoing examples of present work on the biology of parasitism should make clear the excitement of imminent discovery and the promise that the research not only will reap new scientific knowledge, but also will enable applications of immense practical consequences. Further study offers tremendous opportunity for progress.

RESEARCH BRIEFING PANEL ON SOLAR-TERRESTRIAL PLASMA PHYSICS

Charles Kennel ***(Chairman),*** *University of California at Los Angeles, Los Angeles, Calif.*

Jonathan Arons, *University of California at Berkeley, Berkeley, Calif.*

Roger Blandford, *California Institute of Technology, Pasadena, Calif.*

Ferdinand Coroniti, *University of California at Los Angeles, Los Angeles, Calif.*

Martin Israel, *Washington University, St. Louis, Mo.*

Louis Lanzerotti, *AT&T Bell Laboratories, Murray Hill, N.J.*

Dennis Papadopoulos, *University of Maryland, College Park, Md.*

Robert Rosner, *Harvard University, Cambridge, Mass.*

Frederick Scarf, *TRW Systems, Redondo Beach, Calif.*

Staff

Donald C. Shapero, ***Staff Director,*** *Board on Physics and Astronomy, National Research Council*

Charles K. Reed, ***Rapporteur,*** *Board on Physics and Astronomy, National Research Council*

Helene Patterson, ***Staff Assistant,*** *Board on Physics and Astronomy, National Research Council*

Allan R. Hoffman, ***Executive Director,*** Committee on Science, *Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON SOLAR-TERRESTRIAL PLASMA PHYSICS

1 INTRODUCTION[1]

Radio waves were first reflected from the earth's ionosphere—the very edge of space—in the 1920s. Plasma waves were discovered in the laboratory at about the same time. Soon thereafter scientists began to develop a new branch of physics—plasma physics*—to answer the questions stimulated by observations such as those above. Why do laboratory gas discharges exhibit unexpected collective behavior? How

[1]This research briefing is based on the report of the Subpanel on Space and Astrophysical Plasma Physics of the *Survey of Physics* (in preparation [see App. B]), and on the report of the Committee on Solar and Space Physics of the Space Science Board entitled *Solar System Space Physics in the 1980's; a Research Strategy* (1980). This briefing was reviewed and endorsed by the Committee on Solar and Space Physics on August 1, 1984.

*Plasma physics combines concepts from electromagnetism, fluid physics, statistical mechanics, and atomic physics into a unified methodology for the study and practical use of the nonlinear collective interactions of charged particles with each other and with electric and magnetic fields. All physical conditions involving thermal energy densities exceeding 1 electron volt per atom must use plasma physics. A fluid description, magnetohydrodynamics, is used to describe the large-scale properties of plasma systems, and plasma kinetic theory is used to describe the microscopic processes that influence the large-scale properties.

do radio waves propagate in the ionosphere? Why does the appearance of sunspots on the sun presage magnetic storms and auroral displays at earth? What roles do magnetic fields play in stars and galaxies? By the late 1940s, physicists had concluded that plasma is a distinct, fourth state of matter, with properties that are unique because cooperative long-range interactions among the charged particles are far more important than collisions. Moreover, it had become clear that, except for occasional cool regions like planets and their atmospheres, most of the universe is in the plasma state.

Modern plasma physics began in the 1950s. The initiation of research aimed at harnessing the energy source of the stars—thermonuclear fusion—and the first launch of an artificial earth satellite catalyzed the development of the technology needed to study hot plasmas in the laboratory and natural plasmas in space. Moreover, the remarkable discoveries of the solar wind and of the Van Allen radiation belts demonstrated that plasma physics is essential to understand the space environment of the earth and its effects on civilian and defense systems operating in this environment.

The two principal applications of contemporary plasma physics are to fusion research and to solar-terrestrial research—the study of the chain of physical processes that starts with the generation of the sun's magnetic field in the solar interior and links it to activity at the sun's visible surface and, ultimately, to the earth's ionosphere and atmosphere (Section 2). Our understanding of the plasma processes occurring in this interaction chain has made rapid progress in the past decade. The increasing precision of measurements, numerical modelling, and theory applied to solar-terrestrial plasma problems amounts to a revolution in technique relative to 10 years ago. The experimental diagnosis and theoretical interpretations of many space plasma processes now match in precision the best of current laboratory practice. As a result, the study of space plasmas has also become one of the most active areas of basic plasma research.

Research on solar-terrestrial plasmas is advancing our theoretical understanding of many basic plasma processes—for example, how convective turbulence generates magnetic fields, how plasma systems adjust their structure and dynamics by the process of magnetic reconnection, and how particles are accelerated to high energies. These and many other space plasma processes are nonlinear, and their study is contributing to a new branch of mathematical physics—nonlinear dynamics—which has recently led to exciting advances in many branches of physical science.

An increasingly clear perception of the functions of the links in the solar-terrestrial interaction chain has led to a design for a unified study of the solar-terrestrial system as a whole—the proposed multi-spacecraft International Solar-Terrestrial Physics Program (Section 3). Together with the Solar Optical Telescope, discussed in Section 2, this program would be the fundamental underpinning of solar-terrestrial plasma research for the next 10 years. We recommend that the United States commit to participating in the International Solar-Terrestrial Physics Program, by initiating funding in FY 1986.

This research briefing panel has the special responsibility to point out new opportunities. In Section 4, we will discuss the growing impact of solar-terrestrial research on astrophysical plasma physics. The plasma phenomena in the solar system have proven to be examples of processes that also occur in other stars, near neutron stars and black holes, and in galaxies. The sun and solar system have become a laboratory in which astrophysical plasma processes can be studied *in situ* and with a precision attainable nowhere else. As a result, space and astrophysical plasma physicists have begun to work closely together over the past 5 years, and a new and broad research field is beginning to develop.

The great importance of numerical computations to all plasma physics is taken up in Section 5. Such computations have already critically advanced research on fusion plasmas, and we are confident they can be equally important to solar-system and astrophysical plasma physics. We recommend early initiation of a program of large-scale numerical computations dedicated to these subjects, because detailed mathematical models clarify fundamental understanding, strengthen the interpretation of observations, and, in the case of the solar-terrestrial interaction chain, might ultimately lead to some predictive capability. Such a program would also encourage organizationally the already fruitful interactions between space and astrophysical plasma research.

2 THE SOLAR-TERRESTRIAL SYSTEM

Research in this century has revealed a chain of interactions, almost all of which involve plasma physics, that connects the generation of the sun's magnetic field, activity at the surface of the sun, the generation of the solar wind, and the dynamic behavior of the earth's magnetosphere and upper atmosphere.

The first link in the interaction chain is the sun's magnetic field.

Historically, the subject of solar-terrestrial physics began with the observation that auroral and magnetic activity at earth is correlated with the 22-year cycle of the solar magnetic field. Today, no problem in solar-terrestrial research presents a greater challenge than understanding the generation and dynamics of the solar magnetic field. How does convective turbulence below the visible surface of the sun create the sun's weak magnetic field and the persistent concentrations of magnetic field that are sunspots? Recent observations of the microstructure of the sun's magnetic field challenge our understanding of solar turbulence: almost all the surface field is concentrated into tiny regions covering only 1/1000 of the surface area. The ultimate goal remains to understand how the remarkably well-ordered solar cycle arises out of such turbulent chaos.

New observational tools are at hand to help answer some of the challenging questions before us. By observing the periodic motions at the surface of the sun, the new technique of solar seismology is now revealing for the first time the structure of the solar convection zone, at the depths within the sun where the solar magnetic field is created. The Solar Optical Telescope, scheduled for a Shuttle flight in 1990, will measure photospheric and chromospheric structures with the extraordinary spatial resolution of 70 km, providing hitherto unobtainable information essential to theories of magnetohydrodynamic turbulence and solar-surface magnetic fields and dynamics. Later, this instrument is scheduled to become the principal component of the Advanced Solar Observatory, to be flown on a continuing basis on an orbiting platform.

The turbulence in the sun's outer layer heats the tenuous plasma in the solar corona to a temperature of several million degrees; however, the exact mechanism by which the plasma is heated remains uncertain. Ten years ago, *Skylab* observations revealed that the magnetic field emerging from the visible surface evolves into a complex set of open and closed coronal structures that persist for several solar rotation periods (27 days) and mysteriously resist destruction by the sun's differential rotation. The closed field regions—coronal loops—appear to channel energy into the corona and keep it there until microscopic plasma processes dissipate it. These regions are the sites of many small flares, and occasionally of a giant flare, in which magnetic energy is converted into heat and energetic particles, perhaps by a magnetic reconnection process. On the other hand, the energy flowing outward on open field lines drives a plasma wind that expands into interplanetary space. This solar wind carries hot coronal plasma and the sun's open magnetic flux past all the planets of the solar system, interacting with each to form a distinctive magnetosphere.

Our next challenge is to increase our understanding of coronal structure and plasma dynamics and of the processes regulating the flow of energy at each level of the sun's atmosphere. We hope ultimately to integrate this understanding into a unified model that accounts for coronal heating, solar flares, and the generation of the solar wind.

The solar wind is a supersonic, strongly ionized flow that carries plasma, energy, angular momentum, and magnetic field throughout the solar system, interacts with the planets and comets, and fills a three-dimensional volume of space called the heliosphere. The heliosphere terminates where the solar wind is decelerated to subsonic speeds by its interactions with interstellar matter. The solar wind is an excellent laboratory for the investigation of plasma processes of importance to basic plasma physics, solar-terrestrial research, and astrophysics. For example, fully developed magnetohydrodynamic turbulence can be conveniently studied in the solar wind but not in the laboratory. The variability of the solar corona creates a variety of structures propagating in the solar wind whose study has enriched our understanding of magnetohydrodynamics. One of these, the collisionless interplanetary shock, which triggers magnetic storms at earth, accelerates energetic particles by processes thought similar to those that accelerate galactic cosmic rays. The cosmic rays from the galaxy diffuse into the solar system by scattering from solar wind turbulence. As a result, the galactic cosmic-ray flux received at earth is largest at the minimum activity phase of the solar cycle.

Our next challenge is to synthesize our growing understanding into quantitative models that relate coronal activity to the three-dimensional behavior of the solar wind and its propagating structures. To do so, we will need to understand the microscopic processes that regulate plasma transport and the acceleration and propagation of energetic particles.

The solar wind is the plasma link between sun and earth. The solar wind does not directly hit the earth, but is largely deflected by the earth's magnetic field; the name magnetosphere refers to the region enclosing the earth in which the geomagnetic field organizes the behavior of the plasma. Within the magnetosphere, the geomagnetic field traps energetic particles to form radiation belts whose intensity is regulated by microscopic plasma turbulence. Solar-wind energy is coupled into the magnetosphere by the process of magnetic reconnection. When it occurs in the magnetopause, the boundary layer separating the solar wind from the magnetosphere, the reconnection opens part of the earth's magnetic field to the solar wind and stretches the open field into a long magnetic tail extending hundreds of earth radii downstream. A compensating reconnection process in the magnetic tail converts magnetic energy into particle energy and generates strong magnetohydrodynamic flows, which are directed away from the earth tailward of the reconnection region and towards the earth's night side earthward of the reconnection region. Although the consequences of reconnection are clear, the plasma physics of reconnection is not well understood. In the magnetopause, reconnection appears to be spatially structured and temporally variable. In the geomagnetic tail reconnection can be either steady or explosive. One thing is certain: the variability of the solar wind stimulates violent, unsteady magnetospheric flows, which lead to events called substorms, whose effects are felt in the auroral ionosphere and throughout the magnetosphere and tail.

The earth's magnetic field couples the flow in the magnetosphere and tail to the ionosphere and upper atmosphere. Nonlinear plasma processes associated with the electrical currents directed along the field lines connecting flowing magnetospheric plasma to the ionosphere accelerate electrons to about 10 keV. These electrons create the aurora when they hit the atmosphere. The same processes also heat ionospheric plasma, which then boils back out into space to modify the magnetohydrodynamic flow. Some scientists have suggested that the solar-terrestrial interaction chain reaches deep into the atmosphere to modify terrestrial weather and climate; however, the mechanisms by which this might occur have not been identified.

Thus far, solar-terrestrial research programs have concentrated on elucidating individual links in the causal chain connecting the solar wind to the magnetosphere and ionosphere. Our challenge today is to increase our understanding of each of the local plasma processes influencing the magnetosphere's structure and dynamics so that we can begin to create quantitative models that start with the solar wind's observed state and calculate the magnetosphere's and ionosphere's responses.

The most spectacular manifestation of the solar-terrestrial interaction chain is the magnetic storm. The first evidence that a large solar flare might occur is the appearance of a complex sunspot group in the sun's photosphere. Prompt electromagnetic radiation arrives at earth a few minutes after the energy in the coronal magnetic fields associated with the sunspot group is suddenly released. Energetic solar-flare protons are guided by the solar wind and magnetospheric magnetic field into the earth's polar atmosphere soon afterward. The enhanced ionospheric plasma produced by the energetic protons attenuates the radio noise received from cosmic radio sources. Several days later, a shock wave passes over earth, enveloping it in dense, hot solar-flare plasma that compresses the magnetosphere. Substorms increase in frequency and strength, expel hot plasma tailward, and inject hot plasma into the earth's equatorial inner magnetosphere to form a "ring current," which creates the geomagnetic field depression and activity that first motivated the name magnetic storm. The aurorae intensify and move to low latitudes, creating a dense highly disturbed ionospheric plasma that interferes with radio communications. Intense wind systems, sometimes of a worldwide scale, are generated in the upper atmosphere.

Many practical systems, both civilian and defense, and all our manned space endeavors must operate in the highly variable and potentially hostile plasma environment of the earth and solar system. Plasma processes in this environment also influence and even disrupt important ground-based systems over local and regional scales. Ground-based high-frequency communication systems in the earth's polar regions can be entirely blacked out by magnetic storms. Spacecraft have been electronically disabled by violent electrical discharges that occur when hot plasma from the magnetic tail envelops the spacecraft. The risk of such disasters can be reduced only by continuing attention to the effects of the plasma environment on spacecraft systems. It is clear that to work in the space environment, we must understand it.

3 THE INTERNATIONAL SOLAR-TERRESTRIAL PHYSICS PROGRAM

There is general agreement within the research community about the objectives of Solar-Terrestrial Research (Appendix A). Furthermore, there is remarkable unanimity that we can now begin the quantitative study of the solar-terrestrial system as a whole. The unanimity is expressed in the proposal currently before the administration to initiate U.S. participation in the International Solar-Terrestrial Physics Program (ISTP). Spacecraft provided by the National Aeronautics and Space Administration (NASA), the European Space Agency (ESA), and the Japanese Institute for Space and Aeronautical Sciences (ISAS) would make simultaneous, coordinated, detailed measurements of key links in the solar-terrestrial chain. The ISTP program would consist of six elements. Two spacecraft would study the sun and solar wind—

ESA's Solar and Heliospheric Observatory (SOHO) and NASA's Wind spacecraft. Three would study critical regions of the earth's magnetosphere identified by previous research—POLAR (NASA), EQUATOR (NASA), and GEOTAIL (ISAS) (their names indicating the regions to be studied). ESA's multiple spacecraft CLUSTER would study small-scale plasma turbulence and boundary layers in the magnetosphere. Modern communications and data processing technology will make it possible for scientific investigators to obtain a unified set of data from all ISTP spacecraft, a prerequisite to quantitative understanding of the solar-terrestrial system as a whole.

The ISTP draws on the experience of previous coordinated programs, such as the International Geophysical Year (1957–58) and the International Magnetospheric Study (1977–80).* The ISTP is a central part of a more comprehensive effort that includes other funded spacecraft, such as the International Solar-Polar Mission (NASA-ESA), the Upper Atmospheric Research Satellite (NASA), and EXOS-D (Japan). In addition, many ground-based facilities, airborne observatories, balloon and rocket campaigns, and other spacecraft can provide data for collaborative research projects. Active plasma experiments, carried out by Shuttle and other means, will enhance, and be enhanced by, the ISTP.

The United States has already provided much of the leadership in conceiving and designing ISTP, and ISTP will enable the United States to secure a position of future leadership. If it is successful, the ISTP data network, and the collaborative mode of research engendered by its existence, might well outlive the original ISTP spacecraft and shape the way that future solar-terrestrial research will be carried out.

Only a coordinated program, such as ISTP, can study the complex interaction between the parts and the whole of the solar-terrestrial system—an interaction characteristic of all large-scale plasma systems. By studying individual links in the interaction chain, ISTP will provide qualitatively new measurements critical to understanding many fundamental plasma processes—such as dynamo magnetic field generation, particle acceleration, and magnetic field reconnection—that are also important to both laboratory and astrophysical plasma research. By studying several links in the chain simultaneously, ISTP will define the contexts in which these processes occur, a prerequisite for quantitative

*Neither the solar-terrestrial interaction chain nor the role of plasma processes in it were well defined at the time of the International Geophysical Year, and the International Magnetospheric Study did not include the sun and solar wind, and so did not attempt to study the whole interaction chain.

understanding. By studying the system as a whole, ISTP will motivate and test a series of increasingly comprehensive models, whose goal is first to understand, and then to predict, the highly variable plasma environment of the sun, solar wind, and earth. More and more practical systems will have to function in this plasma environment. For all the above reasons,

We recommend: that the United States initiate its participation in the International Solar Terrestrial Physics Program as scheduled in FY 1986.

4 THE IMPACT OF SOLAR-TERRESTRIAL PLASMA PHYSICS ON ASTROPHYSICS

Recent events have reaffirmed the essential unity of laboratory, solar-system, and astrophysical plasma physics that was perceived by the pioneers of the subject 50 years ago. The three branches of plasma physics are unified by a shared preoccupation with a common set of basic physical problems (Appendix B). Yet, during the past 25 years, each branch developed more or less independently, as each individually explored the opportunities of technique and application inspired by modern laboratory, space, and computational technology. A new trend is beginning to emerge. The realization is growing that solar-terrestrial research has developed problems in plasma physics that not only are fundamental to basic plasma physics but are broadly applicable to many astrophysical systems. In particular, a new field of space and astrophysical plasma physics is developing that takes as its starting point the great advances in all three branches of plasma physics during the past 25 years.

The first successes of this new, broader discipline are already in evidence. Terrestrial magnetospheric physics provided a coherent framework for understanding the magnetospheres of Mercury, Jupiter, and Saturn. The upcoming spacecraft encounters with Uranus and Neptune will further test and extend magnetospheric concepts. Since Venus is unmagnetized, studies of Venus' interaction with the solar wind will provide a foundation for investigating the plasma physics of cometary magnetospheres, which soon will be studied by spacecraft for the first time.

Magnetic fields thread their way among planets, stars, galaxies, and probably even clusters and superclusters of galaxies. By transporting angular momentum, turbulent magnetic fields may regulate the accretion flow onto white dwarfs, neutron stars, and black holes. The gravitational energy released when accreting matter falls onto such compact objects may power the enormous energy outputs of galactic x-ray sources and quasars. The strong convective turbulence and buoyancy forces that develop in any gravitationally bound atmosphere will rapidly dissipate and expel magnetic fields. Hence, in nearly all astrophysical bodies, new magnetic fields must be continuously regenerated by turbulent magnetic dynamo processes. The sun provides the only accessible laboratory in which the dynamo mechanism in a highly conducting plasma can be observed and our theoretical dynamo models tested.

Plasma outflows like the solar wind are found throughout the cosmos. The *Einstein* spacecraft made the startling discovery that most stars have hot x-ray emitting coronae, implying that convective envelopes, magnetic dynamos, surface magnetic activity, and strong winds are common. Much of the interstellar medium is filled with hot, low-density plasma from blended stellar winds and supernova remnants. Since the densities, velocities, temperatures, and magnetic field strengths in the interstellar plasma are similar to those of the solar wind, solar system measurements are automatically relevant to the physics of the interstellar medium. Surrounding gas pressure collimates the winds from newly forming stars into bipolar jets; similar jets are observed in the exotic compact object, SS-433, and in radio galaxies. Relativistic winds, possibly consisting of electrons and positrons, are thought to flow away from pulsars and supermassive black holes in active galactic nuclei. The basic physics that underlies all these astrophysical winds was formulated and observationally confirmed in studies of the solar wind.

Super-energetic plasmas occur throughout the astrophysical universe. Quasars and active galactic nuclei convert prodigious amounts of energy into the relativistic plasmas that form the galactic scale radio jets. Trillion-volt electric potentials develop near the magnetic polar regions of pulsars. Radiation from the particles accelerated in these potentials is thought to trigger a pair-production cascade that fills the surrounding magnetosphere with an electron-positron plasma and powers a highly relativistic wind. In our galaxy, supernova shocks accelerate the cosmic rays up to a total energy density that is comparable with the thermal and magnetic energy densities of the interstellar medium. The plasma processes responsible for particle acceleration have long been studied in the solar-terrestrial context—in planetary magnetospheres, at planetary bow shocks and interplanetary shocks, and in solar flares. The acceleration models derived from these observations guide and test the underlying plasma physics of their astrophysical counterparts. For example, although it is not possible to detect the plasma processes responsible for particle acceleration in supernova shocks, theories of cosmic ray acceleration can be tested by measurements of solar system shocks together with measurements of galactic cosmic ray energy spectra and composition.

The challenge for the next decade is becoming clear: space and astrophysical plasma physicists must work side by side to formulate quantitative models of solar system and astrophysical phenomena in which magnetohydrodynamics and plasma physics are essential for the interpretation of observations, recognizing that most of the universe is in the plasma state.

5 LARGE-SCALE NUMERICAL MODELS

Plasma physics was a pioneer in the successful utilization of large-scale computations for fluid, magnetohydrodynamic, hybrid, and kinetic models. Most of the progress since the late 1960s has been in fusion research and nuclear weapons phenomenology. The establishment of a computational facility dedicated to magnetic fusion significantly advanced understanding of magnetic confinement systems and of fusion and basic plasma processes. The establishment, in 1979, of NASA's highly successful Solar-Terrestrial Theory Program made numerical models and simulations regular tools in solar system plasma research and has prepared that research community for the next, more advanced stage.

In the past decade, our research on solar-terrestrial plasmas has achieved a measure of quantitative understanding through the use of numerical models. The study of plasmas beyond the solar system has developed more slowly than space plasma physics for a fundamental reason: the microscopic plasma processes that regulate the behavior of astrophysical systems cannot be observed directly, as can those in the solar system and in the laboratory. Now, however, the modern theoretical and computational techniques developed to understand laboratory and space measurements have opened the door to modelling of the plasmas in the still larger and more exotic environments of astrophysics. Since the underlying astrophysical plasma processes cannot be directly detected, we believe that the best strategy will be to create numerical models at the large-scale system level that postulate small-scale plasma and radiation processes and to iterate between the system and process levels until quantitative agreement with observation is achieved. It is our perception that, for the first time, the current level of development of numerical technology, theory, and observations gives such a strategy a significant chance of success.

The continued development of numerical technology will advance many branches of science. We foresee that many problems of a scale that requires today's national computing facilities will soon be addressable by local university and laboratory facilities. This will only increase the importance of numerical modelling to our subjects. Nonetheless, we believe that the leading research on many solar system and astrophysical plasma models will continue to be done on the most advanced computing facilities existing at any given time, because these models involve a complex interplay between large- and small-scale processes.

Thus far, the responsibility for the maintenance and advancement of state-of-the-art computing facilities has been a national one, because it is beyond the capability of single institutions and because a national scope provides an adequate pool of users. America's existing advanced computational facilities, devoted to defense, fusion research, and meterology, have been used on a piecemeal basis for solar system and astrophysical plasma research. These busy facilities do not have space physics and astrophysics as an institutional objective, and researchers in these fields must make individual agreements to secure access to advanced computing. In some cases, American researchers have had to journey to Europe or Japan in order to perform large-scale computations.

To enhance the research of the International Solar-Terrestrial Physics Program, and to promote fruitful interchanges between solar system and astrophysical plasma physics,

We recommend: a national computational program dedicated to basic plasma physics, space physics, and astrophysics, which will provide and maintain state-of-the-art technology appropriate to large-scale theoretical models and simulations. Such a program should ensure ready access to advanced computing on the basis of peer review.

Many problems central to solar-system and astrophysical plasma research are ready for advanced numerical modelling. If these problems are combined with others in hydrodynamics and general astrophysics that should be included in the program, the increased scientific demand, and more importantly, the scientific payoff, would justify the dedicated effort we propose. The success of numerical computations applied to laboratory plasma problems gives us confidence that a computational program would significantly advance space and astrophysical plasma research. Solar-terrestrial research would be the principal initial contributor to, and beneficiary of, such a program. Solar-terrestrial models that successfully meet the test of detailed measurements of both large- and small-scale processes would substantially increase our confidence in models of more distant astrophysical systems.

APPENDIX A
THE GOALS OF SOLAR-TERRESTRIAL PHYSICS

The research goals of solar-terrestrial physics have been succinctly defined by the Space Science Board in *Solar-System Space Physics, a Research Strategy* (National Academy of Sciences, 1980).

Two basic principles guided the formulation of these goals:

- The objectives of solar-system space research are to understand the physics of the sun; the heliosphere; and the magnetosphere, ionosphere, and upper atmosphere of the earth, other planets, and comets.
- Studies of the interactive processes that generate solar variation and link it to the earth should be emphasized, because they reveal basic physical mechanisms and have useful applications.

To understand better all the processes linking the solar interior to the corona, we need to study the following:

- The sun's global circulation, how it reflects interior dynamics, could modify luminosity, and is related to the solar cycle.
- The interactions of solar plasma with strong magnetic fields-active regions, sunspots, and fine-scale magnetic knots—and how they cause the release of solar-flare energy to the heliosphere.
- The solar corona's energy sources and the physics of its large-scale weak magnetic field.

To understand better the transport of energy, momentum, energetic particles, plasma, and magnetic field through interplanetary space, we need to study the following:

- First and foremost, the coronal processes that govern the generation, structure, and variability of the solar wind.
- The three-dimensional properties of the solar wind and heliosphere.
- The plasma processes that regulate solar-wind transport and accelerate energetic particles throughout the heliosphere.

To understand better the time-dependent interaction between the solar wind and earth, we need to study the following:

- The transport of energy, momentum, plasma, and magnetic and electric fields across the magnetopause.
- The storage and release of energy in the earth's magnetic tail.
- The origin and fate of the plasma(s) within the magnetosphere.
- How the earth's magnetosphere, ionosphere, and atmosphere interact.

To understand better the entire upper atmosphere as one dynamic, radiating, and chemically active fluid, we should study the following:

- The radiant energy balance, chemistry, and dynamics of the mesosphere and stratosphere and their interactions with atmospheric layers above and below.
- The worldwide effects of the magnetosphere's interaction with the polar thermosphere and mesosphere and the role of electric fields in the earth's atmosphere and space environment.
- The effects of variable photon and energetic particle fluxes on the thermosphere and on chemically active minor constituents of the mesosphere and stratosphere.

To understand better the effects of the solar cycle, solar activity, and solar-wind disturbances upon earth, we need to:

- Provide to the extent possible simultaneous measurements on many links in the chain of interactions linking solar perturbations to their terrestrial response.
- Create and test increasingly comprehensive quantitative models of these processes.

To clarify the possible solar-terrestrial influence on earth's weather and climate, we need to:

- Determine if variations in solar luminosity and spectral irradiance sufficient to modify weather and climate exist.
- Ascertain whether any processes involving solar and magnetospheric variability can cause measurable changes in the earth's lower atmosphere.
- Strengthen correlation studies of solar-terrestrial, climatological, and meteorological data.

APPENDIX B
THE UNITY OF LABORATORY, SOLAR SYSTEM, AND ASTROPHYSICAL PLASMA PHYSICS

Although the experimental techniques we use to study laboratory, solar system, and astrophysical plasmas differ, our research on them is unified by the use of plasma concepts. The Committee on Space Physics (National Academy of Sciences, 1984) and the Subpanel on Space and Astrophysical Plasma Physics of the *Survey of Physics* (National Academy of Sciences, in preparation) have concluded that these disciplines share a common set of basic physical problems. These include:

1. magnetic field reconnection;
2. the interaction of turbulence with magnetic fields;
3. the behavior of large-scale plasma flows and their interactions with magnetic and gravitational fields;
4. the acceleration of energetic particles;
5. particle confinement and transport;
6. collisionless shocks;
7. beam-plasma interactions and the generation of electromagnetic radiation; and
8. collective interactions between neutral gases and plasmas.

The fact that such problems emerge from a variety of contexts demonstrates their general significance and suggests that their solution will find further applicability in contexts we cannot imagine today. The existence of such general problems provides a basis upon which a network of common interests, personal interactions, and a common discipline are being built.

RESEARCH BRIEFING PANEL ON SELECTED OPPORTUNITIES IN PHYSICS

Hans Frauenfelder ***(Co-chairman),*** *University of Illinois, Urbana, Ill.*

Mildred S. Dresselhaus ***(Co-chairman),*** *Massachusetts Institute of Technology, Cambridge, Mass.*

William F. Brinkman, *Sandia National Laboratory, Albuquerque, N. Mex.*

Praveen Chaudhari, *International Business Machines Corporation, Yorktown Heights, N.Y.*

Joseph Cerny, *Lawrence Berkeley Laboratory, Berkeley, Calif.*

John M. Dawson, *University of California at Los Angeles, Los Angeles, Calif.*

Harold P. Furth, *Princeton University, Princeton, N.J.*

Daniel Kleppner, *Massachusetts Institute of Technology, Cambridge, Mass.*

Alexei A. Maradudin, *University of California at Irvine, Irvine, Calif.*

*The panel gratefully acknowledges assistance in preparing its report from Gordon Baym, Robert B. Palmer, Bruce Partridge, David Schramm, Lee S. Schroeder, Andrew M. Sessler, Charles P. Slichter, and Robert Wagoner. Thanks are also due to participants in a planning meeting to select the topics treated in the Physics Research Briefing; those not named above include Richard Brewer, IBM; Dale Compton, Ford Motor Company; Robert Park, University of Maryland and American Physical Society/Washington; and Robert Richardson, Cornell University.

Francis M. Pipkin, *Harvard University, Cambridge, Mass.*

John Schiffer, *Argonne National Laboratory, Argonne, Ill.*

Watt W. Webb, *Cornell University, Ithaca, N.Y.*

David Wilkinson, *Princeton University, Princeton, N.J.*

Staff

Donald C. Shapero, ***Staff Director,*** *Board on Physics and Astronomy, National Research Council*

Ellen Williams, ***Rapporteur,*** *University of Maryland, College Park, Md.*

Helene Patterson, ***Staff Assistant,*** *Board on Physics and Astronomy, National Research Council*

Florence Wong, ***Secretary,*** *Board on Physics and Astronomy, National Research Council*

Allan R. Hoffman, ***Executive Director,*** *Committee on Science, Engineering, and Public Policy*

REPORT OF THE RESEARCH BRIEFING PANEL ON SELECTED OPPORTUNITIES IN PHYSICS

INTRODUCTION

Physics has been rich in discoveries of fundamental laws of nature. These discoveries have influenced other fields of science from mathematics to biology and medicine and have spawned entire industries. Maxwell's electromagnetic theory, for example, forms the basis for understanding electromagnetic phenomena from waves to plasmas and underlies radio, television, radar, and our industrial power networks and modern communication systems. Quantum mechanics provides a framework for portraying physical reality and underlies all of the natural sciences. It has made possible the invention of the transistor, many of the amazing devices of modern solid state electronics, and the laser. Lasers, in turn, are having a major impact on many areas of science. They are also playing a rapidly expanding role in medicine and broad areas of technology, including communications, manufacturing, and national defense.

The discovery of x rays led to the largest single advance in the history of medical diagnostics. The recently created NMR body imaging technique may mark a comparable advance. From the study of electrical noise in the atmosphere, radioastronomy was born, which in turn provided a dazzling new portrait of the universe. Investigations of the properties of nuclei led to the creation of the nuclear power industry. Numerous other examples could be cited. The benefits to society from basic research in physics have been incalculable.

Recent fundamental advances in physics demonstrate that physics is still in a golden age and has never been more vigorous or more productively interactive with other fields. Progress in physics over the last decade has been remarkable. Puzzles that seemed to present insuperable challenges at the beginning of the 1970s have yielded to powerful and elegant theoretical and experimental techniques. The new insights and accomplishments have not only brought greater unity to the various branches of physics but have also strengthened the ties of physics to other areas of science and opened a vast array of new opportunities. Every part of physics has participated in the advance, as the forthcoming report of the Physics Survey Committee will make clear.

The present report focuses on six areas of special opportunity in physics. In order to find areas in which incremental funding may lead to major advances, a planning group (consisting of the Briefing Panel and others listed above) was convened. From twenty suggestions, the following six were chosen: (1) physics at the laser-atomic frontier, (2) relativistic plasma waves, (3) physical properties of deliberately-structured materials, (4) biomolecular dynamics and intercellular cooperativity, (5) cosmology, and (6) nuclear matter under extreme conditions. Large national facilities and well-established programs were explicitly excluded.

The six areas chosen promise to yield fundamental results of great interest. Many of the areas are likely to advance technology and to produce results with impact on the nation's industries; many can be expected to contribute to national security.

All six areas cut across lines of narrow specialization and link different fields. Progress in any one area can thus be expected to stimulate progress in other fields. Cosmology, for instance, connects physics to astronomy and astrophysics. Biological physics interacts strongly with chemistry, biochemistry, biology, medicine, and pharmacology. Deliberately-structured materials relate fundamental aspects of physics to many fields of technology.

1. PHYSICS AT THE LASER-ATOMIC FRONTIER

The laser continues to have a revolutionary impact in the field from which it emerged—atomic physics. The laser-atomic frontier comprises research opportunities generated by joining laser methods with new techniques in atomic and molecular physics.

Opportunities at the laser-atomic frontier include research with trapped particles; femtosecond (10^{-15} sec) spectroscopy and other new forms of spectroscopy; the study of previously inaccessible atomic, molecular, and ionic species; and the development of novel light sources. Advances in these areas will lead to new tests of basic theory and deepen our understanding of the structure of matter at the atomic and molecular level, including the transfer of energy and the nature of chemical reactions. In addition to high scientific interest, many of these topics are likely to yield technological advances. The opportunities selected include the following.

TRAPPED PARTICLES

Using laser light, it is possible to cool to the millikelvin region ions confined in electromagnetic traps and also to cool and trap neutral atoms. In conjunction with recently developed ultrastable tunable lasers, these developments open the way to major advances in high-precision measurements: frequencies to parts in 10^{16}, masses to parts in 10^{11}, new tests of the isotropy of space, and studies of collective motion in plasmas and gases. Applications include advanced atomic clocks, optical frequency standards, metrology, and communications.

NEW SPECTROSCOPIES

Breakthroughs in the generation of femtosecond light pulses make it possible to take "snapshots" of atoms as they collide or molecules as they react, and to observe fast adsorption and relaxation on surfaces. Femtosecond spectroscopy has applications to fast electron circuitry and high-speed instrumentation. Another new type of spectroscopy is carried out with relativistic particle beams. Using the Doppler effect to shift highly stable laser radiation from the visible into the ultraviolet opens the way to new areas of high-resolution spectroscopy of atoms, ions, and molecules. Scientific objectives include studies of relativistic and quantum electrodynamic effects in highly charged ions, high resolution ultraviolet spectroscopy of atoms and molecules, and new types of photoejection measurements.

THE CREATION OF NEW SPECIES

Lasers can produce new species as well as species that have been inaccessible, such as multiply excited atoms, molecular ions, and clusters. Problems such as correlated electron motion, the evolution of matter from single atoms to the condensed state, and catalysis can be investigated. There are many applications to other sciences and to industrial problems. Studies of ionic species can lead to the design of more efficient combustion engines; catalysis is of enormous importance to chemical processing.

MATTER IN INTENSE FIELDS

Laser methods open the way to the study of matter under extraordinary conditions, such as in intense electric and magnetic fields and powerful radiation fields where nonlinear phenomena and multiphoton processes occur. Studies of multiphoton processes made possible by high-power lasers are expected to reveal new aspects of the interaction of photons with matter, and they may open the way to novel types of photochemical processing and isotope separation.

NEW LIGHT SOURCES

The creation of the excimer laser and the generation of extreme ultraviolet light in supersonic atomic beams are two recent examples of light sources from research at the laser-atomic frontier. Other new sources can be expected, with wavelengths from the infrared to the soft x-ray region. Such sources would be valuable to wide areas of atomic physics, chemistry, condensed matter physics, and biology. Using coherent generation of pulsed laser light, for example, it may be possible to provide very short pulses in the soft x-ray region. This technique may provide a laboratory-based alternative to synchrotron radiation light sources for many applications.

2. RELATIVISTIC PLASMA WAVES

Recently an exciting area of plasma physics has emerged—that of relativistic plasma waves in which the electrons or both electrons and ions have relativistic velocities. Exploration of relativistic plasma waves is expected to elucidate recently discovered exotic astrophysical objects and the acceleration of particles to create cosmic rays. An understanding of relativistic collective electromagnetic phenomena may lead to novel particle accelerators and radiation generators. We sketch below some of the specific ideas that need investigation.

PARTICLE ACCELERATORS

The next generation of high-energy accelerators is pushing existing technology close to the limit. The long-range future of particle and nuclear physics may well depend on the development of new acceleration techniques. In addition, new accelerators in the range of 10 MeV to 10 GeV may find application in physics, other sciences, industry, and medicine.

Plasma Accelerators

Space-charge waves in plasmas can produce electric fields that are orders of magnitude larger than can be generated by any other means, with the possible exception of a laser beam focused in a vacuum. These fields can be larger than the fields that bind electrons to atoms (10^9 volts/cm). Such fields can only be maintained in a plasma because they instantly ionize ordinary materials. In a beat-wave accelerator, space-charge waves are generated by sending through the plasma two intense laser beams with a frequency difference equal to the plasma frequency. The combination of high electric fields and phase velocities close to that of light makes such waves good candidates for the acceleration of particles to high energy.

Grating Accelerators

Another novel method uses a grating of droplets illuminated by a powerful short-pulse laser. The resulting plasma grating generates fields estimated to be of several hundred million volts per cm.

Free-Electron Laser Accelerators

Another possibility for constructing an accelerator is to operate a free electron laser "backwards" in an accelerating mode. A further possibility is to operate a free electron laser "forward" and use the resulting radiation to accelerate particles in a conventional linear accelerator. Both of these concepts pose interesting problems and will require developments in current technology.

Laser Development

All of these accelerating schemes require terawatt lasers with good optical quality and very short pulses (shorter than 1 picosecond). Progress with these approaches could be accelerated by a laser development program aimed toward these requirements.

GENERATION OF ELECTROMAGNETIC RADIATION

The inverse of particle acceleration—radiation by energetic particles in plasmas—is a challenging subject of scientific and potentially practical interest. Bursts of radiation that accompany solar activity and are also emitted from a wide variety of astrophysical objects are poorly understood. Recent observations of radiation from pulsars show that the pulses have γ-ray energies up to 10^{15} eV. Such energies imply that the acceleration process occurs over relatively short distances (of the order of 10^3 km) with accelerating fields larger than 10^7 volts/cm. Relativistic waves may be involved. Closer to home, bursts of energetic particles are generated by solar flares; here again strong electric fields must exist in the plasma. The theoretical and computer modeling tools are now reaching the point where progress in understanding these processes can be made.

In the laboratory, copious radiation, roughly up to the thirtieth harmonic of the plasma frequency, has been observed when relativistic electron beams are injected into plasmas. The generation of electromagnetic radiation by means of free-electron lasers is actively pursued at a number of laboratories. Within the last year, free-electron lasers have started operation—two in the microwave, three in the infrared, and one in the visible range. Further configurations especially suited to different parts of the spectrum could be developed. Moreover, theory predicts that energetic electrons streaming through plasmas with strong density fluctuations should yield radiation similar to that produced in present free electron lasers where rippled magnetic fields are used instead of density ripples. At present there are virtually no controlled experiments to investigate these phenomena. The computational tools to address these problems exist, but the resources for using these tools are generally not available to university researchers.

3. DELIBERATELY-STRUCTURED MATERIALS

Condensed-matter science has a long and continuing history of contributions to basic understanding of physical phenomena. General concepts from this field are bringing greater intellectual unity to physics. Ideas from condensed-matter physics and many-body theory are playing a role in such diverse fields as nuclear physics, astrophysics, and theoretical methods of particle physics. While condensed-matter physics is intellectually a rich and fertile field of inquiry for the basic questions of science, it is also playing a vital role in advanced technology from microelectronics and communications to energy systems.

From the vast range of activities in condensed-matter physics, we emphasize only one area of special opportunity: deliberately-structured materials. By deliberately-structured materials, we mean materials designed by man to have particular structures. Their fabrication may involve controlling on an atomic length scale the compositional or structural arrangements of atoms in one or two dimensions. The resulting materials possess remarkable properties not otherwise found in nature—properties that, in many instances, present major scientific puzzles.

Not only are these materials likely to raise fundamental questions of physics, but their study will also reinforce the technological base of the nation's most important device-related industries. Conversely, technologies that have led to advances in the fabrication of very-large-scale integrated circuits can be harnessed to create deliberately-structured materials.

Three research areas that are currently of great scientific interest and that are especially likely to advance through the use of deliberately-structured materials are: (1) studies of surfaces and interfaces between specifically designed microscopic structures; (2) investigations of "disorder" in matter that might be understood more deeply through tailoring of its composition or geometrical structure; and (3) studies of novel collective phenomena in "exotic materials."

SURFACES, INTERFACES, AND THIN FILMS

Beginning with an atomically clean and crystallographically defined surface, it is possible to prepare a two-dimensional film only 1 atom thick by deposition of a second material. The film can be atomically smooth or contain crystallographic discontinuities, depending on the structure of the substrate. Its structure can be further modified by pattern etching on a scale finer than 100 Å. Deposition of additional layers of dopants, or of metals, semiconductors, superconductors, or insulators, leads to atomically sharp interfaces of defined structure and composition. The physics of these low-dimensional systems is fundamentally interesting. Furthermore, studies of the electronic, vibrational, magnetic, and transport properties of interfaces are applicable to the technological needs of the computer and energy industries.

DISORDER IN MATTER

There are many approaches to creating materials with controlled disorder. One method is the random dispersion of very small particles of one material in a matrix of a second. Another is the fabrication of alternating amorphous layers, resulting in two-dimensional disorder within the plane of the layers but one-dimensional order perpendicular to the plane due to the stacking of the layers. Spin glasses present a totally different form of disorder in which the atoms may be compositionally ordered but the spin-spin interactions are random. Metastable (i.e., nonequilibrium) disordered structures represent a fourth class of deliberately-disordered materials. All these materials have properties that are intrinsically interesting and important. In addition, insight into the nature of controlled disorder may provide a basis for understanding technologically important disordered materials such as polymers, glasses, and composite materials and may even lead to a better understanding of problems in computation, optimization, and neurobiology.

EXOTIC MATERIALS

The structure of materials can be arranged in many different ways to produce unique physical properties. For instance, it is possible to deliberately construct ordered crystals in arrays of alternating layers called superlattices. The properties of superlattices can be modified in desirable ways by tailoring the compositional and impurity profiles of the constituent layers. Many of the traditional methods of condensed-matter theory must be extended to account for the small size or the lower dimensionality of such systems. New phenomena, undreamt of a short time ago, have been discovered in deliberately-structured materials. Examples include quantization and fractional quantization of the Hall effect. The same class of materials that is actively used to investigate the quantum Hall effect is also being examined by researchers worldwide for new applications in the computer and communication industries.

NEEDS OF THE FIELD

The overwhelming characteristic of this field is its dependence on a strong infrastructure to support a range of activities from materials preparation and characterization to study of the nature and causes of unusual materials properties. The infrastructure, involving skilled technicians and complex, state-of-the-art apparatus and instruments, does not exist in most universities; and the institutional mechanisms to put the structure in place at universities are often lacking. The intellectual concepts are there to be exploited if the university infrastructure can be enhanced.

4. BIOMOLECULAR DYNAMICS AND INTERCELLULAR COOPERATIVITY

The magnificent complexity of life that we study as biology ultimately reflects underlying physical principles. At each level—biomolecules, supramolecular structures, cells, and organisms—the behavior of the system is governed by physical laws. The detailed description of biological phenomena in terms of physical laws promises to yield rich dividends both in terms of understanding life more deeply and in exploring the physics of complex systems. Research on biological systems connects physics to chemistry, biochemistry, biology, and, ultimately, pharmacology and medicine. The systems under investigation are, of course, the same in all these fields, but the approaches to them are different. The physicist selects the simplest available system, performs quantitative measurements, constructs a model, and attempts to describe the observed behavior fully in terms of known physical laws. The following examples, at different levels of complexity, indicate where this approach is beginning to bear fruit.

BIOMOLECULAR DYNAMICS

Proteins and nucleic acids, the building blocks of life, were called "aperiodic crystals" by Schrödinger. Seen by standard x-ray diffraction, they indeed appear to be like crystals with linear dimensions of a few nanometers. In recent years, however, a far more exciting picture has emerged. Biomolecules are dynamic systems that represent a state of matter different from solids, liquids, and gases. They show many similarities to glasses and spin glasses but are far more sophisticated. The sophistication is not surprising; each biomolecule has undergone more than 10^9 years of research and development and is built for a particular function. A protein, for instance, must have precise mechanical, electrical, thermal, and chemical properties. These properties are not spatially homogeneous but rather vary throughout the protein volume. Through genetic engineering, proteins with atomic modifications in desired places can be produced. The entire field of physical studies of biomolecules is at its very beginning, but the importance and the potential of such studies are increasingly evident. Biomolecules are excellent laboratories for the investigation of basic physical laws. We can study elementary excitations, motions, transport phenomena, and elementary reaction steps. Biomolecular physics overlaps the physics of glasses and spin glasses and complements the chemistry of reactivity, catalysis, and life processes. With increasing knowledge of the relation between structure and function, it will become possible to design and synthesize biomolecules with desired properties for uses in biological research, medicine, and pharmacology. It is also likely that an understanding of the relation between structure and function of biomolecules, together with genetic engineering, will produce spectacular results in other areas. Biosensors have the potential of improving equipment from medicine to information transfer. New biomaterials, such as biomolecular catalysts and artificial photosynthetic systems, may find uses in areas ranging from agriculture to space.

TRANSMEMBRANE SIGNALLING

Transmembrane signalling is conspicuous in brain, nerve, and muscle tissue. Its molecular basis is now directly accessible through a new biophysical technology: recording of the stochastic electrical conductances of individual molecular channels isolated in very small patches of cell membranes or reconstituted into model membranes. Switching of transmembrane currents through molecular channels is controlled by neurotransmitters, messenger molecules, or membrane potential changes. Hundreds of distinguishable membrane channels and nearly as many neurotransmitters have been discovered. How are channel protein conductances switched by binding of a neurotransmitter? How is ion selectivity built into the protein structure? These profound questions and the biological behavior and disorders associated with them appear susceptible to solution with the help of a combination of gene manipulation and biophysical measurements.

MOLECULAR BASIS OF INFORMATION STORAGE

Elegant experiments on the organization of elementary learning and memory in simple animals at the cellular and molecular level imply a variety of basic signalling processes. The molecular basis of memory is, however, still not understood. The combination of advanced biophysical techniques, such as sensitive optical and electron microscopy probes, with hybridoma techniques for producing monoclonal antibodies, offers hope of elucidating the molecular features of information storage and signal transmission. Prospects are good for understanding the molecular mechanisms of auditory transduction and tone discrimination and the mechanism of visual signal transmission from the photoreceptor. A new family of sensors may develop from these biophysical experiments with some hope of revolutionizing the tools of clinical medicine.

INTERCELLULAR COOPERATIVE PROCESSES

Recent developments at the interface between physics and mathematics have generated new analytical concepts dealing with discrete and nonlinear systems. These systems often display regimes of solution characterized by solitons, limit cycles, disorder, and chaos. These concepts represent many intercellular cooperative processes and have provided understanding of several systems, such as the cooperative heart muscle cell contraction in the heart beat and its failure in ventricular fibrillation.

Physicists have generated new conceptual models of distributed memory. These models, while speculative at present, may, if successful, find ultimate applications in robotics and computer architecture. The theory reproduces many of the foibles and perhaps some of the strengths of the human memory.

5. COSMOLOGY

Few areas of science enjoy as wide an appeal to scientists and laypersons as cosmology. Fascination with the origin and fate of the universe is deeply imbedded in the human intellect, and speculations about these ultimate questions have stimulated some of the most profound thinking in modern science. As in all other scientific disciplines, the basis of cosmology lies in experimental observations. Accurate measurements of the abundances of light elements, for example, fit the specific predictions of a "hot big-bang" model for our universe. Light nuclei were formed when the universe was a few minutes old and had a temperature of 10^9 K—a fantastic extrapolation of known physics. The same hot big-bang model predicted the 3 K cosmic radiation a decade before its discovery.

COSMOLOGY AND PARTICLE PHYSICS

Though uniquely successful in explaining these fundamental observations, the hot big-bang model by itself fails to explain others.

- Why is our universe so asymmetric in matter and antimatter?
- Why is the ratio of baryons to photons in the universe $\sim 3 \times 10^{-10}$?
- Why is the kinetic energy in universal expansion so nearly equal to the universe's gravitational potential energy at the current epoch?
- How did the universe reach the same temperature to a part in 10^4 at points not causally connected?

Two recent ideas from particle physics may answer these fundamental questions and still fit into the basic hot big-bang picture. Grand Unified Theories (GUTs) when applied in the very early universe (age $\sim 10^{-35}$ sec) supply answers to the first two questions. Thus, laboratory tests of GUTs (e.g., by means of proton-decay experiments) may also resolve important cosmological problems. The last two questions above may be answered by a model in which the universe expanded exponentially for a brief time early in its history. This "inflation" is supposed to have been driven by the vacuum energy of a scalar field perhaps related to the Higgs particle that is required by the new theories of electromagnetic and weak interactions. We can, of course, turn the arguments around. Already cosmological data are constraining and guiding GUTs, making the early universe effectively a particle physics laboratory.

WHAT LIES AHEAD?

The long-awaited evidence to describe the fate of the universe could come from either of two lines of attack. One approach attempts to measure directly the deceleration of the universe over the past few billion years. The other approach looks for "invisible" mass density that might eventually reverse the current universal expansion, leading to a hot recollapse. Massive neutrinos, axions, and other exotic particles are under active study in this context by particle physicists and cosmologists. Primordial nucleosynthesis has recently advanced a strong argument against the existence of a large component of dark matter in baryonic form; thus, the search for an exotic mass component has intensified.

Another area of current excitement is the origin and development of large-scale structure—galaxies, clusters of galaxies, etc. The structure seen today is assumed to grow from density perturbations in the early universe. However, observational limits on fluctuations in the thermal background radiation at small angular scale severely constrain models for the origin of structure. Most models that lead to the observed matter clumping also predict fluctuations of the radiation exceeding observed limits. The standard picture must be modified and extended.

Cosmology is at present ripe with opportunity and certain to be full of surprises. The combination of the space program with major advances in astronomical instrumentation is rapidly pushing our horizons deeper into the universe across a wide range of the electromagnetic spectrum. Theory, drawing on all areas of physics, is creating a plausible and testable cosmological model rich with beautiful physics and unanswered questions.

6. NUCLEAR MATTER UNDER EXTREME CONDITIONS

Until recently the study of nuclear matter was restricted to normal values of density and low temperature. With the acceleration of heavy nuclei to relativistic energies, an important new dimension has been added to these studies, namely, the ability to deliver unprecedented amounts of energy and momentum into the volume of the colliding nuclei, a volume large relative to that of a nucleon. With this tool a new era for studying nuclear matter opens up; for the first time we can test the response of such matter to extreme conditions of energy density and compression.

SCIENTIFIC OPPORTUNITIES

Central nucleus-nucleus collisions at high energies open up new avenues that cross the traditional boundaries of particle, nuclear, and astrophysics. The broad goals of such a program—to explore the cooperative behavior in extended hadronic systems with many interacting degrees of freedom—are complementary to those of particle physics in the same sense that condensed-matter physics, with its varied and often unexpected phenomena, transcends the underlying atomic physics. The salient difference is that nuclear studies can themselves shed light on the incompletely understood mechanism of confinement in the basic quantum chromodynamic theory.

Foremost among the opportunities is the possibility of producing a new state of matter called the quark-gluon plasma. Through numerical lattice-gauge calculations, quantum chromodynamics predicts that confined hadronic matter, in which quarks are bound together to form baryons and mesons within the nucleus, will cease to exist at sufficiently high values of compression or energy density. A new phase of matter appears in which the quarks and gluons are freed from the confines of the hadrons. Current theoretical estimates and extrapolations of existing data indicate that the phase transition can be explored in at least two distinct domains accessible in heavy-ion collisions. The domain of *maximum nuclear compression* is a baryon-rich environment expected to occur at energies of a few GeV/nucleon in the center-of-mass frame. In this domain, the heaviest nuclei can still stop each other completely. Densities up to 10 times that of normal nuclei should occur; such densities may be achieved through enhancement of existing accelerators. The domain of *high-energy density* is expected to occur in the central rapidity region left after the colliding nuclei pass through each other at energies of several tens of GeV/nucleon in the center-of-mass frame. Exploration of this domain, expected to be meson-rich with energy densities in excess of 10 times those of ordinary nuclei, awaits construction of future accelerators.

Colliding very heavy nuclei at relativistic energies allows one to study new states of matter and to map out the equation of state of highly compressed/excited nuclear matter, while providing the first indications of the dynamical behavior of extended regions of matter under the most extreme baryon and energy densities accessible in the laboratory.

Suitable facilities will soon be available for accelerating light ions to the 15-225 GeV/nucleon range. These new capabilities are in the region expected to yield maximum nuclear compression and to provide an important extension of our studies of nuclear matter under extreme conditions.

By their very nature, these complex experiments present a major challenge to experimentalists. Detector requirements are similar to those found in particle physics but with a high degree of detector segmentation and enhanced particle identification because of the large numbers of particles expected. Research and development efforts on these detectors are needed so that we will be adequately prepared to measure the complex event structures expected in these interactions.

In nuclear theory, substantial efforts are required to produce theoretical benchmarks for comparison with experimental results. Examples include large Monte Carlo intranuclear cascade calculations and lattice-gauge calculations with inclusion of finite quark masses and finite baryon densities.

IMPACT ON OTHER AREAS

In addition to potential technological spinoffs from the development of new detectors, research on central high-energy nucleus-nucleus collisions links nuclear physics to other fields. Information gained about the transition from the quark-gluon plasma to hadronic matter is important to cosmologists for understanding how the present state and distribution of matter in the universe evolved. Data from the regime of maximum compression are required to understand the state of matter in the deep interior of neutron stars. There is no other way such information on the high-density equation of state and other properties of matter vital for condensed-matter astrophysics can be obtained in the laboratory. These data will facilitate the interpretation of observations from future orbiting x-ray telescopes such as the AXAF and will help to unravel questions about the birth and evolution of neutron stars. Particle physicists have been studying the short-range behavior of quantum chromodynamics and processes of quark confinement; high-energy nuclear collisions will allow us for the first time to investigate the long-range behavior of this theory and the processes of deconfinement.

CONCLUSIONS

The six areas selected in the present report comprise only a small part of physics. But they exhibit clearly the characteristics of physics—enormous diversity, the search for fundamental laws, strong connections to many other sciences, and technological and industrial applications. Each of the six areas is at a stage where incremental funding may produce major progress. The reasons for the high leverage vary from one area to another, but in each area, there could be rapid progress through meeting a need either for seed funding to promote new research or for new instrumentation, especially medium-cost instrumentation for university research laboratories. Several of the subjects—relativistic plasma accelerators, cosmology, and parts of biophysics—do not have established mechanisms for support and review.

The new areas within nuclear physics, biophysics, and cosmology described above are each in their infancy; exploiting the opportunities to do new forefront physics addressing fundamental questions about matter, life, and the universe could be greatly accelerated by a modest investment of seed money.

Each of the new areas within atomic physics, plasma physics, and condensed-matter physics described above has a need for medium-cost instrumentation that is difficult for university laboratories to obtain within existing funding frameworks. An increment of funding targeted to the instrumentation problem in these areas could unleash untapped scientific productivity in areas with potential applications to technology and national security.