The Future of Supercomputing: An Interim Report

Committee on the Future of Supercomputing, National Research Council

ISBN: 0-309-52675-2, 58 pages, 8.5 x 11, (2003)

This free PDF was downloaded from:
http://www.nap.edu/catalog/10784.html

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

THE FUTURE OF SUPERCOMPUTING

An Interim Report

Committee on the Future of Supercomputing

Computer Science and Telecommunications Board

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS

Washington, D.C.

www.nap.edu

Printed in the United States of America

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON THE FUTURE OF SUPERCOMPUTING

SUSAN L. GRAHAM, University of California, Berkeley, *Co-chair*
MARC SNIR, University of Illinois at Urbana-Champaign, *Co-chair*
WILLIAM J. DALLY, Stanford University
JAMES DEMMEL, University of California, Berkeley
JACK J. DONGARRA, University of Tennessee, Knoxville
KENNETH S. FLAMM, University of Texas at Austin
MARY JANE IRWIN, Pennsylvania State University
CHARLES KOELBEL, Rice University
BUTLER W. LAMPSON, Microsoft Corporation
ROBERT LUCAS, University of Southern California
PAUL C. MESSINA, Argonne National Laboratory (part-time)
JEFFREY PERLOFF, University of California, Berkeley
WILLIAM H. PRESS, Los Alamos National Laboratory
ALBERT J. SEMTNER, Naval Postgraduate School
SCOTT STERN, Kellogg School of Management, Northwestern University
SHANKAR SUBRAMANIAM, University of California, San Diego
LAWRENCE C. TARBELL, JR., Eagle Alliance
STEVEN J. WALLACH, Chiaro Networks

Staff

CYNTHIA A. PATTERSON, Study Director and Program Officer
PHIL HILLIARD, Research Associate
MARGARET MARSH HUYNH, Senior Project Assistant

COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD

Preface

High-performance computing is important in solving many kinds of complex problems in domains from weather science and biology to national security. U.S. government spending on supercomputing has been relatively flat over the last 10 years and has declined compared to industrial or commercial purchases of high-performance systems.[1] Some observers associate the trends with the end of the Cold War, during which both national security and scientific research needs were believed to justify spending on a range of high-performance computing programs. Others point to the influence of the changing marketplace for computing systems overall, which has seen an increase in demand for less expensive systems of less than maximal performance.

Several factors have led to the recent interest in reexamining the rationale for federal investment in research and development in support of high-performance computing, including (1) continuing changes in various component technologies and their markets, (2) the evolution of the computing market (and particularly the high-end supercomputing segment), (3) experience with several systems using the clustered processor architecture, and (4) the evolution of the problems, many of them mission-driven, for which supercomputers are used.

The Department of Energy's (DOE's) Office of Science expressed an interest in sponsoring a study by the Computer Science and Telecommunications Board (CSTB) of the National Research Council that would assess the state of U.S. supercomputing capabilities and relevant research and development. Spurred by the development of the Japanese vector-based Earth Simulator supercomputer, the Senate's Energy and Water Development Appropriations Committee directed the Advanced Simulation and Computing (ASC) Program of the National Nuclear Security Administration (NNSA) at DOE to commission a study (in collaboration with DOE's Office of Science) by the National Research Council. Congress also commissioned a study by the JASONs[2] to identify the distinct requirements of the stockpile stewardship program and its relation to the ASC acquisition strategy.

CSTB convened the Committee on the Future of Supercomputing to assess prospects for supercomputing technology research and development in support of U.S. needs, to examine key elements of context—the history of supercomputing, the erosion of research investment, the changing nature of problems demanding supercomputing, and the needs of government agencies for supercomputing capabilities—and to assess options for progress. The committee has been tasked with preparing two

---

[1]Debra Goldfarb, IDC's HPC Industry Analyst, presentation to the committee on May 23, 2003.

[2]The JASONs, formed in 1959, are a select group of scientific advisors who consult with the federal government chiefly on classified research issues. For more on the JASONs, see Ron Southwick, 2002, Elite Panel of Academics Wins Fight to Continue Advising Military, *The Chronicle of Higher Education*, June 7. Available at <www.globalsecurity.org/org/news/2002/020607-jason.htm>.

reports—a brief interim report and a final in-depth report. This interim report is intended only to establish context—the history and current state of supercomputing, application requirements, technology evolution, the socioeconomic context, and so on—and to identify some of the issues that may be explored in more depth in the second phase of the study. In order to provide feedback as soon as possible, this interim report has been developed on a very tight time line. It is based on committee deliberations and briefings received from numerous experts at two committee meetings.[3] It does not contain formal conclusions or recommendations. The committee expects that its understanding of some background issues will change as it collects more data and deepens its analysis in the final report, anticipated in late 2004.

The committee thanks the many individuals who contributed to its work. The people who briefed the committee at one of the plenary meetings are listed in Appendix C. Their willingness to answer our questions was most helpful. Stanford University, with the able local support of Pamela Elliott and Charles M. Orgish, hosted the second plenary meeting of the committee. The sponsors of the report at the Department of Energy—Daniel Hitchcock, Fred Johnson, José L. Muñoz, Dimitri Kusnezov, and Hans Ruppel—have been most supportive and responsive in helping the committee to do its work. The reviewers of the draft report provided insightful and constructive comments that contributed significantly to the clarity of the report.

The work of the committee was made considerably easier because of the participation of excellent NRC staff members. Marjory Blumenthal, the outgoing director of CSTB, shared her wisdom and knowledge with the committee until she left the NRC in June. She will be greatly missed by all of us who have worked with her on this study and in earlier CSTB activities. Jon Eisenberg, Herb Lin, and Richard Rowberg have given valuable counsel to the chairs. Margaret Marsh Huynh is providing first-rate administrative and logistical support to the committee. Liz Fikre, the report editor, has taught us much about clear exposition. Phil Hilliard has willingly and competently given research support to the committee. Finally, Cynthia Patterson, the study director, has been an outstanding partner and mentor to the chairs. Her contributions have strengthened both the study and the interim report.

Susan L. Graham and Marc Snir, *Co-chairs*
Committee on the Future of Supercomputing

---

[3]The speakers are listed in Appendix C.

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's (NRC's) Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Roy Radner, New York University,
Ahmed H. Sameh, Purdue University,
Charles L. Seitz, Myricom, Inc.,
Allan Snavely, San Diego Supercomputing Center,
Francis Sullivan, IDA Center for Computing Sciences, and
Paul R. Woodward, University of Minnesota.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Samuel H. Fuller, Analog Devices, Inc. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Contents

xi

Executive Summary

By definition, supercomputers are those hardware and software computing systems that provide close to the best currently achievable sustained performance. The performance of supercomputers, which is normally achieved by introducing parallelism, is typically much better than that of the vast majority of installed systems. Supercomputers are used to solve complex problems, including the simulation and modeling of physical phenomena such as climate change, explosions, or the behavior of molecules; the analysis of data from sources such as national security intelligence, genome sequencing, or astronomical observations; or the intricate design of engineered products. Their use is important for national security and defense, as well as for research and development in science and engineering.

As the uses of computing have increased and broadened, supercomputing has become less dominant than it once was. Many interesting applications require only modest amounts of computing, by today's standards. Because of the increase in computer processing power, many problems whose solution once required supercomputers can now be solved on relatively inexpensive desktop systems. That change has caused the computer industry, the research and development community, and some government agencies to reduce their attention to supercomputing. Yet problems remain whose computational demands for scaling and timeliness stress even our current supercomputers. Many of those problems are fundamental to the government's ability to address important national issues. One notable example is the Department of Energy's computational requirements for nuclear stockpile stewardship.

The government has sponsored studies of a variety of supercomputing topics over the years. Some of those studies are summarized in the body of this report. Recently, questions have been raised about the best ways for the government to ensure that its supercomputing needs will continue to be satisfied in terms of both capability and cost-effectiveness. To answer those questions, the National Research Council's Computer Science and Telecommunications Board convened the Committee on the Future of Supercomputing to conduct a 2-year study to assess the state of supercomputing in the United States and to give recommendations for government policy to meet future needs. This study is sponsored jointly by the Department of Energy's Office of Science and by its Advanced Simulation and Computing (ASC) Program.

This interim report, presented approximately 6 months after the start of the study, reflects the committee's current understanding of the state of U.S. supercomputing today, the needs of the future, and the factors that contribute to meeting those needs. After such a short time, the committee is not yet ready to comment in detail on the specifics of existing supercomputing programs or to present specific findings and well-supported recommendations. Although the committee has made considerable progress in understanding the current state of supercomputing and how it got there, it still has much more work to do before it develops recommendations.

1

Many technical, economic, and policy issues need to be addressed in this study. They include (1) the computational needs of present and future applications and approaches to satisfying them; (2) the balancing of commodity components and custom design in supercomputer architectures and the effects of software design improvements and industry directions on that balance; (3) the interplay of research, development, prototyping, and production in creating innovative advances; (4) the extent and nature of direct government involvement to ensure that its needs are met; and (5) the important requirement that the present not be neglected while the future is being determined. Although this report touches on each of those topics, the committee has not completed its consideration of them.

The particular technical approaches of any program that develops or uses supercomputing represent a complex compromise between conflicting requirements and an assessment of risks and opportunities entailed in various approaches. An evaluation of these approaches requires a detailed understanding of (1) the relevant applications, (2) the algorithms used to solve those application problems, (3) the performance likely to be achieved by codes that implement these algorithms on different platforms, (4) the coding efforts required by various approaches, (5) the likely evolution of supercomputing technology over multiple years under various scenarios, and (6) the costs, probabilities, and risks associated with different approaches. In its final report, the committee will seek to characterize broadly the requirements of different application classes and to examine the architecture, software, algorithm, and cost challenges and trade-offs associated with these application classes—keeping in mind the needs of the nuclear stockpile stewardship program, the broad science community, and the national security community. (Note that a separate, classified report by the JASONs is expected to identify the distinct requirements of the stockpile stewardship program and its relation to the ASC acquisition strategy.) The committee believes that it would be unwise to significantly redirect or reorient current supercomputing programs before careful scientific consideration has been given to the issues described above. Such changes might be hard to reverse, might reduce flexibility, and might increase costs in the future.

In the period ahead, the committee will continue to learn and to analyze. A workshop focused on applications, to be held in the fall of 2003, will include a number of applications experts from outside the committee. Its purpose will be to identify both the computational requirements of important applications and the opportunities to adapt and evolve current solutions so as to benefit from advances in algorithms, architectures, and software. In addition, the committee will meet with experts who are developing solutions for applications of particular importance for national defense and security within the Department of Energy (DOE) and the National Security Agency (NSA). The committee will also meet with managers of supercomputing facilities, procurement experts, industrial supercomputer suppliers, experts on computing markets and economics, and others whose expertise will help to inform it.

SUPERCOMPUTING TODAY

According to the June 2003 list of the 500 most powerful computer systems in the world, the United States leads the world in the manufacture and use of supercomputers, followed by Japan.[1] A number of other countries make or use supercomputers, but to a much lesser degree.

Virtually all supercomputers are constructed by connecting large numbers of compute nodes, each having one or more processors and a common memory, by an interconnect network (a switch). Supercomputer architectures differ in the design of their nodes, their switches, and the node-switch interfaces. Higher node performance is achieved by using commercial scalar microprocessors with 64-bit data paths intended primarily for commercial servers or nodes designed specially for supercomputing, rather than the high-volume, 32-bit scalar microprocessors used in workstations and lower capability cluster systems. The custom nodes tend to use special mechanisms such as vectors or multithreading to reduce memory latency rather than relying solely on the more limited latency avoidance afforded by caches. High-bandwidth, scalable interconnects are typically custom-built and more expensive than the

---

[1]The TOP500 list is available at <www.top500.org>.

more widespread Ethernet interconnects. Custom switch use is often augmented by custom node/switch interfaces.

The highest-ranked system in the TOP500 list is the Japanese-built Earth Simulator, released in the spring of 2002 and designed specifically to support geoscience applications. That system has custom multiprocessor vector-based nodes and a custom interconnect. The emergence of that system has been fueling recent concerns about continued U.S. leadership in supercomputing.

The system software that is used on most contemporary supercomputers is some variant of Unix, either open source or proprietary. Programs are written in FORTRAN, C, and C++ and use a few standard application libraries. All of these supercomputers use implementations of the message passing interface (MPI) standard to support message-passing-style internode communication. Relatively little standardization exists for other aspects of software environments and tools.

EVOLUTION IN SUPERCOMPUTING

A major policy issue for supercomputing is the proper balance between investments that exploit and evolve current supercomputing architectures and software (the evolutionary aspect) and investments in alternative approaches that may lead to a paradigm shift (the innovative aspect).   Both aspects are important.

At this stage in the study, the committee sees the following advantages for an evolutionary approach to investment and acquisition. First, much useful work is getting done using the existing systems, and their natural successors can be expected to continue that work. In addition, there are no obvious near-term architectural alternatives: The promising technology breakthroughs, such as processor-in-memory, streaming architectures, and the like, that might revolutionize supercomputing are far off in the future and less than certain. Of course, higher capability would enable better solutions to be obtained faster. But history suggests that even when revolutionary advances come along, they do not immediately supplant existing architectures. Different problems benefit from different architectures and no one design is universally best. The committee sees a need for evolutionary investments in all major approaches to supercomputing that are currently pursued: clusters built entirely of commodity components; scalable systems that reuse commodity microprocessors together with custom technology in the interconnect or the interconnect interface; and systems in which microprocessors, the interconnect, and their interface are all customized.

Although some advantages also accrue from evolutions in software, the committee sees a need for investment that would accelerate that evolution. The advantages of commodity architectural components might be more easily realized if more applications were redesigned, better custom software was provided, and the cost of bringing software to maturity was better appreciated. The benefit of software investment is that it tends to have continuing value as architectures evolve. At the same time, both the maintenance and the evolution of legacy applications must be anticipated and supported.

Finally, the committee observes that uncertainties in policy and inconsistencies over time can be both disruptive and expensive. Unexpected pauses in an acquisition plan or failure to maintain a diversity of suppliers and products can cause both the suppliers and the skilled workforce to divert their attention from supercomputing. It can be difficult and expensive to recover the supply of expertise.

INNOVATION FOR SUPERCOMPUTING

Innovation in supercomputing stems from application-motivated research that leads to experimentation and prototyping, to advanced development and testbeds, and to deployment and products. All the stages along that path need continuous, sustained investment in order that the needs of the future will be met. If basic research activities are not supported, revolutionary advances are less

likely; if experimentation and advanced development are not done, the promising approaches never ripen into products.

In supercomputing, innovation is important in architecture, in software, in algorithms, and in application strategies and solution methods. The coupling of these aspects is equally important.

Major architecture challenges stem from the uneven performance scaling of different components. In particular, as the gap between processor speeds, memory bandwidth, and memory and network latency increases, new ideas are needed to increase bandwidth and hide (tolerate) latency. Additionally, as new mechanisms are introduced to address those issues, there is a need for ways to supply a stable software interface that facilitates exploiting hardware performance improvements while hiding the changes in mechanism.

The need for software innovation is motivated by its role as an intermediary between the application (the problem being addressed) and the architectural platform. Innovation is needed in the ways that system software manages the use of hardware resources, such as network communication. New approaches are needed for ways in which the applications programmer can express parallelism at a level high enough to reflect the application solution and without platform-specific details.   Novel tools are needed to help application-level software designers reason about their solutions at a more abstract and problem-specific level. Software technology is also needed to lessen future dependence on legacy codes. Enough must be invested in the creation of advanced tool and environment support for new language approaches so that users can more readily adopt new software technology. Importantly, advances in algorithms can sometimes improve performance much more than architectural and software advances do.

More realistic simulations and modeling require not only increased supercomputer performance but also new methods to handle finer spatial resolution, larger time scales, and very large amounts of observational or experimental data. Additional applications challenges for which innovation is needed are the coupling of multiple physical systems, such as the ocean and the atmosphere, and the synthesizing of a physical system's design by analytic estimation of its properties. Emerging applications in areas such as bioinformatics, biological modeling, and nanoscience and technology are providing both new opportunities and new challenges.

THE ROLE OF THE GOVERNMENT

There are several important arguments for government involvement in the advancement of supercomputers and their applications. The first is that unique supercomputing technologies are needed to perform essential government missions and to ensure that critical national security requirements are met. Furthermore, without the government's involvement, market forces are unlikely to drive sufficient innovation in supercomputing, because the innovators—like innovators in many other high-technology areas—do not capture the full value of their innovations. Historically, it seems that innovations in supercomputing have played an important role in the evolution of today's mainstream computers and have provided important benefits by virtue of their use in science and engineering. These benefits seem to significantly exceed the value captured by the initial inventors.

It appears that the ability of government to affect the supercomputing industry has diminished because supercomputing is a smaller fraction of the total computer market and computer technology is increasingly a commodity. This situation requires a careful assessment of the most effective ways for government to influence the future of supercomputing.

1

Introduction

Supercomputers are systems that provide significantly greater sustained performance than is available from contemporary mainstream computer systems. In applications such as analysis of intelligence data, weather prediction, or climate projection, supercomputers enable the generation of information that would not otherwise be available or that could not be generated in time to be actionable. Supercomputing can also accelerate scientific research in important areas, such as physics, biology, or medicine. It can augment experimentation or replace it with simulation, thus reducing the cost and increasing the accuracy and repeatability of experimentation in science and engineering. Further, supercomputing has the potential to suggest entirely novel experiments that can revolutionize our perspective of the world. It enables faster evaluation of design alternatives, thus improving the quality of engineered products. The value of supercomputers derives from the problems they solve, not from the innovative technology they showcase. The technology must be motivated by the application requirements.

Historically, better performance was achieved using faster logic and more parallelism—that is, by performing many computations and data accesses concurrently. Today, although there is some promising device research, parallelism is still the primary approach.

The performance of supercomputers should be measured in terms of the time required to solve problems of interest. Some problems, such as searches for patterns in data, can be broken down into subproblems that can be solved independently and the results easily combined later. Thus, a collection of PCs that are intermittently available and are connected by a low-speed network such as the Internet can exhibit supercomputing performance, as shown by the example of SETI@Home.[1] For such problems, a computational grid can replace a conventional supercomputer.[2] However, many important problems requiring high-performance computing, such as the modeling of fluid flows, do not admit that kind of decomposition. While these problems can be solved using parallelism, dependencies among the subproblems necessitate frequent exchange of data and partial results, requiring significantly better communication (higher bandwidth and lower latency) between the computation and data storage loci than that achieved by network-connected PCs.[3]

---

[1]"SETI@home: the Search for Extraterrestrial Intelligence." Available at <setiathome.ssl.berkeley.edu>.

[2]A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities (I. Foster and C. Kesselman, 1999, Computational Grids, *The Grid: Blueprint for a New Computing Infrastructure*, Morgan-Kaufman.).

[3]Although the grid does not replace supercomputers for many high-end applications, grid computing enhances our ability to solve the kinds of problems for which large amounts of computation and often-unique data are essential. In addition, the grid provides the infrastructure within which tightly coupled supercomputers reside. It enables efficient access to remote or specialized computation resources and efficient exchange of results between collaborating scientists.

5

This report focuses mostly on the latter kind of problem and, hence, on "one machine room" systems. To achieve high performance on problems of interest, such supercomputers need not only the ability to perform operations at a high rate but also support for high-bandwidth, low-latency internal communication, large memories, and high-performance I/O subsystems. They also need suitable software, both high-quality systems software—such as compilers and operating systems—and well-tuned applications software.

To maintain focus, this report does not address networking (i.e., the external communication requirements of supercomputers) except to note its importance, nor does the report address special-purpose systems such as signal processors.

STUDY CONTEXT

Much has changed since the 1980s, when a variety of agencies invested in developing and using supercomputers and when the High Performance Computing and Communications Initiative (HPCCI), which bridged and built on these agency efforts, was conceived—and the 1990s, when the HPCCI evolved into a broader and more diffuse program of computer science research support.[4] More recently, federally supported high-performance computing research has deemphasized computer architecture research and begun to emphasize the networked grid for high-performance computing, an emerging industry interest,[5] as a unifying concept. Whereas early investments in high-performance computing research were shown to have had trickle-down benefits for mainstream computing,[6] recent trends cloud the picture for such benefits. Meanwhile, there is increasing evidence of and concern about technical leadership in Japan, where the industry has benefited from sustained government investment.[7] Concern about the diminishing U.S. ability to meet national security needs led to a recommendation in 2000 that DOD subsidize a Cray computer development program as well as invest in relevant long-term research.[8]

CSTB convened the Committee on the Future of Supercomputing, sponsored jointly by the DOE Office of Science and DOE's Advanced Simulation and Computing (ASC) Program to assess the state of supercomputing in the United States, including the characteristics of relevant systems and architecture research in government, industry, and academia and the characteristics of the relevant market. Specific questions of interest to both the sponsors and Congress are listed in Box 1.1.

---

[4]The proliferation of PCs and the rise of the Internet commanded attention and resources, diverting attention and effort from research in high-end computing. There were, however, efforts into the 1990s to support traditional high-performance computing. See, for example, NSF, 1993, *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing*. NSF Blue Ribbon Panel on High Performance Computing. Arlington, Va.: National Science Foundation, August.

[5]Barnaby J. Feder. 2000. Supercomputing Takes Yet Another Turn. *The New York Times*, November 20, p. C4.

[6]Computer Science and Telecommunications Board (CSTB), National Research Council. 1995. *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Infrastructure*. Washington, D.C.: National Academy Press. This report noted the time-machine quality of high-performance systems.

[7]Japanese support for indigenous capabilities has led to U.S. allegations of dumping, the most recent of which were resolved by an arrangement featuring U.S. resale of Japanese machines by a U.S. vendor (see William M. Bulkeley, 2001, Outlook improves for U.S. supercomputer access, *The Wall Street Journal*, March 2, p. B6.) Meanwhile, U.S. vendors have long complained about controls on high-performance computer exports (see Ted Bridis, 2001, Study suggests easing limits to export high-performance computers overseas, *The Wall Street Journal*, June 8, p. B5.)

[8]Defense Science Board. 2000. *Report of the Defense Science Board Task Force on DOD Supercomputing Needs*. Washington, D.C.: Office of the Under Secretary of Defense for Acquisition and Technology, October 11.

Box 1.1 Concerns of the Sponsor and of Congress

Issues of concern to the Department of Energy:

- How should the nation approach research and development of the highest end computers to ensure its future leadership in science and technology?
- What is the economic model (e.g., commercial investment, particle accelerator, or submarine) for high-performance computers?
- How do we allocate the investment between scientific applications, mathematical algorithms, system software, hardware architectures, hardware engineering, and so on?
- What is the current state of the art of supercomputing in the United States and the rest of the world?
- What/who are the requirements drivers of supercomputing?
- What are supercomputing's "gold nuggets" and how might they be exploited?
- Does (or should) open source software have a role to play?
- What are the shortfalls and how do we address them over the next 3, 5, 10, and 20 years? What are the costs of the solutions?
- What should be the U.S. supercomputing vision? What role should (or can) the government play?

Questions of particular interest to the Senate's Energy and Water Development Appropriations Committee, which funds the Department of Energy:

- What are the mission requirements driving the ASC program?
- How much capacity is needed, and when is it needed over the next 10 years?
- What is the maximum capability required in the top ASC platform, and when over the next 10 years?
- Was the NNSA wise to abandon custom-designed chips and vector architecture for much cheaper commodity-chip-based, massively parallel systems?
- What level of customization is needed for the various government interests in supercomputing—for example, weapons design, molecule modeling and simulation, cryptanalysis, bioinformatics, climate modeling?
- How effective are the current or planned ASC platforms in addressing the requirements of the program?
- Are there alternative architectures, interconnect technologies, systems software and tools, or other approaches that will improve the performance of future ASC platforms?
- If so, can industry supply the required alternative architectures and software? That is, is industry properly motivated to do this? Or must government fundamentally lead the development of these alternatives?
- Is the current ASC approach the most cost-effective and efficient manner of achieving the desired capability and capacity?
- Finally, as they relate to the ASC mission requirements, what are the costs and benefits of investing more heavily in capacity now and deferring acquisition of capability machines, so as to take advantage of the falling price per teraflop?

SOURCE: Daniel Hitchcock, DOE, José L. Muñoz, DOE, and Clay Sell, Senate Energy and Water Development Appropriations Committee, presentations to the committee, March 6, 2003. Many of the questions in the second list are to be addressed by the JASONs' study, mentioned elsewhere in this report.

ABOUT THIS INTERIM REPORT

This interim report focuses on stage setting and context—for example, the history and current state of supercomputing and the socioeconomic context—together with some identification of issues being addressed by the committee. The committee expects that its understanding of these issues will change as it collects more data and deepens its analysis for the final report. The short time it had to prepare the interim report did not allow the committee to develop findings or recommendations. The presentations to the committee and the collective knowledge and experience of committee members have, however, enabled it to come to a preliminary understanding of some important issues, which are outlined in this interim report. However, the report does not document the detailed evidence that supports these views. The final report will provide the needed depth of information and analysis. Since the committee appreciates the desire to use this interim report to inform budgetary discussions for FY 2005, it shares some of its initial views in this preliminary form. Chapter 2 outlines the history and current state of supercomputing. The importance of continuity is summarized in Chapter 3. Chapter 4 discusses the need for research and innovation. Chapter 5 addresses the role of the government in ensuring the future health of supercomputing.

2

Supercomputing Past and Present

This chapter provides background material on supercomputing to establish key elements of context. A summary of reports and government activities illuminates the recent history of supercomputing. A brief overview of the current state of supercomputing technology follows.

PREVIOUS REPORTS AND RECENT FEDERAL INITIATIVES

During the past few decades, a number of reports have dealt with supercomputing and its role in science and engineering research. The first of the modern reports is the *Report of the Panel on Large Scale Computing in Science and Engineering* (the Lax report).[1] The Lax report made four basic recommendations: (1) increase access for the science and engineering research community to regularly upgraded supercomputing facilities via high bandwidth networks, (2) increase research in computational mathematics, software, and algorithms necessary to the effective and efficient use of supercomputing systems, (3) train people in scientific computing, and (4) invest in research and development basic to the design and implementation of new supercomputing systems of substantially increased capability and capacity, beyond that likely to arise from commercial requirements alone. In 1985, following the guidelines of the Lax report, the National Science Foundation (NSF) established five supercomputer centers.

Following the renewal of four of the five NSF supercomputer centers in 1990 and the possible implications for them contained in the 1991 High Performance Computing Act (P.L. 102-194), the National Science Board (NSB) commissioned the NSF Blue Ribbon Panel on High Performance Computing to investigate the future changes in the overall scientific environment due to rapid advances in computers and scientific computing.[2] The panel's report, *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing* (the Branscomb report), recommended a significant expansion in NSF investments, including accelerating progress in high-performance computing through computer and computational science research.

In 1995, NSF formed a task force to advise it on the review and management of the supercomputer centers program. The chief finding of the *Report of the Task Force on the Future of the NSF*

---

[1]Panel on Large Scale Computing in Science and Engineering. 1982. *Report.* Sponsored by the Department of Defense and the National Science Foundation in cooperation with the Department of Energy and the National Aeronautics and Space Administration. Washington, D.C., December 26.

[2]National Science Foundation. 1993. *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing*. NSF Blue Ribbon Panel on High Performance Computing, August.

9

Supercomputer Centers Program (the Hayes report)[3] was that the Advanced Scientific Computing Centers funded by NSF had enabled important research in computational science and engineering and had also changed the way that computational science and engineering contribute to advances in fundamental research across many areas. The recommendation of the task force was to continue to maintain a strong Advanced Scientific Computing Centers program.

Congress asked the National Research Council's Computer Science and Telecommunications Board to examine the High Performance Computing and Communications Initiative (HPCCI).[4] CSTB's 1995 report *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Infrastructure* (the Brooks/Sutherland report)[5] recommended the continuation of government support of research in information technology; the continuation of the HPCCI; funding of a strong experimental research program in software and algorithms for parallel computing machines; HPCCI support for precompetitive research in computer architecture (but end direct HPCCI funding for development of commercial hardware by computer vendors and for "industrial stimulus" purchases of hardware); and the development of a teraflop computer as a research direction rather than a destination.

In 1997, following the guidelines of the Hayes report, NSF established two Partnerships for Advanced Computational Infrastructure (PACI), one with the San Diego Supercomputer Center as a leading-edge site and the other with the National Center for Supercomputing Applications as a leading-edge site. Each partnership includes participants from other academic, industry, and government sites. A third participant, the Pittsburgh Supercomputer Center, was added in 2000. The PACI program is scheduled to end in the fall of 2004.

In 1999, the President's Information Technology Advisory Committee's (PITAC's) *Report to the President: Information Technology Research: Investing in Our Future* (the PITAC report) made recommendations similar to those of the Lax, Hayes, and Branscomb reports.[6] PITAC found that federal information technology R&D is too heavily focused on near-term problems and that investment was inadequate. The committee's main recommendation was to create a strategic initiative to support long-term research in fundamental issues in computing, information, and communications.

Supercomputing applications have also been studied. In 1999, *The Biomedical Information Science and Technology Initiative* found that because the number of biomedical researchers who could profit from using supercomputing facilities was increasing, the National Institutes of Health (NIH) should take a strong leadership position and help support the national supercomputer centers.[7]

The 2003 report *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (the Atkins report)[8]

---

[3]National Science Foundation. 1995. *Report of the Task Force on the Future of the NSF Supercomputer Centers Program*. September 15.

[4]HPCCI was formally created when Congress passed the High-Performance Computing Act of 1991 (P.L. 102-194), which authorized a 5-year program in high-performance computing and communications. The goal of the HPCCI was to "accelerate the development of future generations of high-performance computers and networks and the use of these resources in the federal government and throughout the American economy" (Federal Coordinating Council for Science, Engineering, and Technology (FCCSET), 1992, *Grand Challenges: High-Performance Computing and Communications*. FY 1992 U.S. Research and Development Program, Office of Science and Technology Policy, Washington D.C.). The initiative broadened from four primary agencies addressing grand challenges such as forecasting severe weather events and aerospace design research to more than 10 agencies addressing national challenges such as electronic commerce and health care.

[5]Computer Science and Telecommunications Board (CSTB), National Research Council. 1995. *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Infrastructure*. Washington, D.C.: National Academy Press.

[6]President's Information Technology Advisory Committee (PITAC). 1999. *Report to the President. Information Technology Research: Investing in Our Future*. February.

[7]National Institutes of Health. 1999. *The Biomedical Information Science and Technology Initiative*. Working Group on Biomedical Computing, Advisory Committee to the Director, National Institutes of Health, June 3.

[8]National Science Foundation. 2003. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. January.

found that scientific and engineering research is pushed by continuing progress in computing, information, and communication technology (among other things) and pulled by the expanding complexity, scope, and scale of today's research challenges. The panel's overall recommendation was that NSF should establish and lead a large-scale interagency and internationally coordinated advanced cyberinfrastructure program (ACP) to create, deploy, and apply cyberinfrastructure in ways that radically empower all scientific and engineering research and allied education. The panel strongly recommended that the U.S. academic research community have access to the most powerful computers that can be built and operated in production mode and that NSF should support five centers that will provide high-end computing resources.

There have also been studies of the use of supercomputing for missions important to the United States such as national security. The DOE Accelerated Strategic Computing Initiative (ASCI) [9] was established in 1995 to transition from a test-based to a simulation-based certification program to analyze and predict the performance, reliability, and safety of nuclear weapons. The first supercomputer, ASCI Red, which had 1 Tflop performance, was delivered in 1996. Other ASCI supercomputers include ASCI Blue, ASCI White, and ASCI Q. The goal of ASCI Purple, scheduled for 2005, is 100 Tflop.

In 1996, a study by the Office of the Director of Defense Research and Engineering (DDR&E) stated that in order for the United States to maintain supremacy in the high-end computing field, a major national security program would be necessary. [10]

Two reports by the General Accounting Office (GAO) examined DOE's use of its computing capabilities. The titles of these reports summarize the GAO findings: *Information Technology: Department of Energy Does Not Effectively Manage Its Supercomputers* [11] and *Nuclear Weapons: DOE Needs to Improve Oversight of the $5 Billion Strategic Computing Initiative*. [12] The first report citied utilization rates that showed, in the GAO's view, that the national laboratories were underutilizing their supercomputing capacity and missing opportunities to share it. (DOE disputed those findings.) The lack of an investment strategy and a defined process was cited as a reason why DOE was not fully justifying its supercomputer acquisitions. The second report found that a lack of comprehensive planning and progress tracking systems in the ASCI program made assessment of the initiative's progress difficult and subjective.

In 1998 NSA and DDR&E joined forces and funding to support the development of the SV-2 (now the X1) by Cray Research in order to meet government needs that could not be met elsewhere in the marketplace. In the third quarter of 2002, Cray delivered five early production versions of the X1. A 1024-processor commercial X1 was delivered in early 2003.

The Department of Defense sponsored the *Report of the Defense Science Board Task Force on DOD Supercomputing Needs*. [13] The task force found that there is a significant need for high-performance computers that provide extremely fast access to very large global memories and that such computers support a crucial national cryptanalysis capability. Task force recommendations included providing additional financial support for the development of the Cray SV-2 (now the X1), developing an integrated system based on commercial off-the-shelf (COTS) microprocessors and a new high-bandwidth memory system, and investing in long-term research on critical technologies.

---

[9] This initiative subsequently became the Advanced Simulation and Computing Program but is still often referenced as ASCI.

[10] Director of Defense Research and Engineering. 1996. *DDRE Integrated Process Team Study— A National Security High End Computing Program*.

[11] General Accounting Office (GAO). 1998. *Information Technology: Department of Energy Does Not Effectively Manage Its Supercomputers*. Report to the Chairman, Committee on the Budget, House of Representatives (GAO/RCED-98-208). Washington, D.C.: GAO, July.

[12] GAO. 1999. *Nuclear Weapons: DOE Needs to Improve Oversight of the $5 Billion Strategic Computing Initiative*. Report to the Chairman, Subcommittee on Military Procurement, House Committee on Armed Services (GAO/RCED-99-195). Washington, D.C.: GAO, July.

[13] Defense Science Board. 2000. *Report of the Defense Science Board Task Force on DOD Supercomputing Needs*. October 11.

In 2001, Charles Holland, principal assistant deputy under secretary of defense for science and technology, and a team of experts authored a report[14] that focused on DOD's research and redevelopment agenda for high-performance computing. The report found that current research does not adequately address medium- to long-term needs and proposed an agenda with three thrusts—technology development, concept demonstration, and industry adoption—to address the challenges of producing innovative ideas and reinvigorating the academic and industry research communities.

Supercomputing architecture was the focus of *Survey and Analysis of the National Security High Performance Computing Architectural Requirements* (the Games report).[15] The survey found that a major investment had been made by the national security community to migrate legacy applications from vector supercomputers to commodity high-performance computers (HPCs). It found that although vector supercomputers process more efficiently than commodity HPCs, most but not all large applications scale well on commodity HPCs. Finally, it reported that some researchers found it increasingly difficult to program distributed-memory commodity HPCs, which had a negative impact on their research productivity. Recommendations were to assess the usefulness of Japanese vector supercomputers, reach out to researchers through the use of OpenMP on shared-memory systems, promote flexibility through software that combines OpenMP and message passing interface and that switches between vector- and cache-based optimizations, and establish a multifaceted R&D program to improve the productivity of high-performance computing for national security applications.

The goal of the DARPA high productivity computing systems (HPCS) program, initiated in 2002, is to provide a new generation of economically viable, high-productivity computing systems for the national security and industrial user community in 2007-2010. It is focused on addressing the gap between the capability needed to meet mission requirements and the current offerings of the commercial marketplace. HPCS has three phases: an industrial concept study currently under way with Cray, SGI, IBM, HP, and Sun; an R&D phase that was awarded to Sun, Cray, and IBM in July 2003 and lasting until 2006; and full-scale development, to be completed by 2010, ideally by the two best vendors from the second phase.

The Defense Appropriations Bill for FY 2002 directed the Secretary of Defense to submit a development and acquisition plan for a comprehensive, long-range, integrated, high-end computing (IHEC) program to Congress by July 1, 2002. The resulting report, *High Performance Computing for the National Security Community,* was released in the spring of 2003. The report recommends an IHEC program that integrates applied research, advanced development, and engineering and prototype development. The applied research element will focus on developing the fundamental concepts in high-end computing and creating a pipeline of new ideas and graduate-level expertise for employment in industry and the national security community. The advanced development element will select and refine innovative technologies and architectures for potential integration into high-end systems. The engineering and prototype development element will build operational prototypes and system level testbeds. The report also emphasizes the importance of high-end computing laboratories that will test system software on dedicated large-scale platforms; support the development of software tools and algorithms; develop and advance benchmarking, modeling, and simulations for system architectures; and conduct detailed technical requirements analysis. The report suggests $390 million per year as the steady-state budget for this program. The program is planned to consolidate existing DARPA, DOE/NNSA, and NSA R&D programs and will feature a joint program office with DDR&E oversight.

In addition to the study by the NRC's Committee on the Future of Supercomputing that resulted in this interim report, two other studies of the future of U.S. supercomputing are under way: one by the National Coordination Office for Information Technology Research and Development (ITRD) and another by the JASONs. The ITRD High-End Computing Revitalization Task Force has been charged with developing a plan and a 5-year roadmap to guide federal investments in high-end computing starting

---

[14]Charles J. Holland. 2001. *DOD Research and Development Agenda for High Productivity Computing Systems*. White paper, June 11.

[15]Richard A. Games. 2001. *Survey and Analysis of the National Security High Performance Computing Architectural Requirements*. June 4.

with fiscal year 2005. The final report is due in August 2003, in time to influence the FY 2005 budget. The JASONs' study, commissioned by DOE at the request of Congress, will identify the distinct requirements of the Stockpile Stewardship Program and its relation to the ASCI acquisition strategy. The JASONs are expected to complete their (classified) report in August 2003.

SUPERCOMPUTING TECHNOLOGY

Vendors

Supercomputers have been manufactured in the United States and abroad since early in the history of the computer industry. Since 1993, a list of the sites operating the 500 most powerful computer systems has been available to the public. This list, called the TOP500, is updated twice a year.[16] Performance is measured by the number of floating point operations performed per second (flops) while executing the LINPACK benchmark to solve a dense system of linear equations.[17]

According to the June 2003 TOP500 list, the United States and Japan dominate the use of and manufacture of high-performance systems (although supercomputers are used in Europe, European computer companies have been limited to the integration of relatively small cluster systems). The TOP500 data show that the United States has a 50 percent share of installed supercomputers, Germany has 11 percent, and Japan has 8 percent, accounting for 69 percent of the total. Another interesting comparison is to look at the aggregate performance by country. From the distribution by performance, the U.S. has 54 percent of the aggregate performance of the TOP500 computers and Japan has 17 percent, together accounting for 71 percent of the total.

Breaking the numbers down by manufacturers, the top three, all U.S. companies, are Hewlett-Packard (32 percent of the TOP500 machines), IBM (31 percent), and SGI (11 percent); together they account for 74 percent of the systems. Performance by manufacturer shows that 35 percent of the performance is attributable to IBM's aggregate share of 31 percent, Hewlett-Packard's 24 percent, and NEC's 12 percent. Ninety-one percent of the top 500 systems are U.S. made.[18]

In summary, in both use and manufacture, the United States is the dominant participant, followed by Japan. A small number of companies dominate the market. Germany is a large user of supercomputing but not a large producer.

Architecture

Contemporary supercomputers are all built by clustering large numbers of compute nodes. They span a spectrum of architectural choices, from clusters that are assembled from low-cost, high-volume components, to systems that are custom built for high-end scientific computing. The main differentiators are the node technology, the switch (sometimes called the interconnect) technology, and the node-switch interface.

---

[16]See <http://www.top500.org/>.

[17]No single number captures system performance across a wide range of applications and architectures. Flops in a dense linear algebra benchmark is but one figure of merit; however, it is the one used for this widely referenced list.

[18]Although these percentages would probably change if different metrics were used, the dominance of the United States over other countries would most likely remain.

Node Technology

Most low-cost clusters use 32-bit Intel or Advanced Micro Devices (AMD) microprocessors. These microprocessors are targeted for low-end servers and are produced in very large volumes (on the order of hundreds of millions). They are not optimized for scientific computing.

Sixty-four-bit microprocessors (Alpha, Power, Sparc, MIPS, Itanium, Opteron) offer the advantages of support for larger memories, a better performing memory subsystem, and support for larger shared-memory multiprocessor (SMP) configurations. The production volumes of these microprocessors are two orders of magnitude smaller than the volumes for 32-bit microprocessors. These microprocessors are mostly targeted for high-end commercial servers, although on occasion vendors will develop SMP configurations that are optimized for scientific computing.

Both 32-bit and 64-bit scalar microprocessors are optimized for single-thread performance on codes that exhibit good temporal and spatial locality.[19] These codes make most of their memory references to an on-chip cache, with good cache reuse. Such processors have limited off-chip bandwidth, can support only a small number (at most 8 or 16) of simultaneously outstanding memory references, and have cache line mechanisms that are not ideal for scientific applications. In high-end application codes that do not make good use of caches, this approach to memory system design leads to a dramatic drop in actual performance when compared with the theoretical peak. This problem is mitigated in processors that employ either multithreading or vectors to generate a large number of outstanding memory references and that therefore tolerate long memory latencies while sustaining high bandwidth (thereby reducing the need for data locality). Over 20 percent of the systems are based on Intel and AMD 32-bit processors. About 8 percent of the systems use vector processors. Approximately 60 percent of the systems use 64-bit processors.

Switch Technology

Low-end clusters, including some on the TOP500 list, use high-volume switched Ethernet technology for the interconnect. Higher bandwidth and lower latency are achieved by using custom interconnects from third-party vendors (e.g., Quadrics and Myricom) or from the system vendors (e.g., Cray, IBM, NEC, and SGI). A key differentiator between systems is the fraction of total system cost allocated to the interconnect: Low-bandwidth networks will represent less than 10 percent of total system cost; a high-bandwidth network may approach half of total system cost. Another important differentiator is the scalability of the interconnect to large numbers of nodes.

Node-Switch Interface

Nodes of low-end clusters connect to the switch via a standard I/O bus, such as peripheral component interconnect (PCI). This choice reuses high-volume, low-cost technology but limits the function and performance of the interconnect, since I/O interfaces are not optimized for fast processor-to-processor communication. In such systems, global bandwidth is an order of magnitude lower than local memory bandwidth. The software for communication typically uses message passing, further increasing communication latency and limiting bandwidth for short messages.

A custom memory-connected interface, typically proprietary, can be used to increase bandwidth, reduce latency, or provide added functionality. Such interfaces are usually paired with higher-performance custom switches.[20] In particular, a custom interface can directly support shared memory communication, allowing a processor to access the memory of a remote node via load and store

---

[19]Temporal locality is the property that data accessed recently in the past are likely to be accessed soon in the future. Spatial locality is the property that data that are stored very near one another tend to be accessed closely in time.

[20]Recently, Intel and other companies have been directly attaching standard interconnects such as Ethernet and Infiniband directly to the memory system rather than via the PCI bus; thus, some of the performance advantages of custom interfaces are becoming available with standard interconnects.

instructions. Since shared memory communication has little software overhead, it has lower communication latency; however, the small number of pending memory references supported by mass-market microprocessors limits global bandwidth. Shared memory support is generally believed to facilitate parallel programming, because of the single name space it provides; it also facilitates the use of a single operating system image to control the entire machine. Approximately half the systems in the TOP500 list use proprietary switch interfaces.

Products

Close to 20 percent of the TOP500 systems are self-made or are assembled by system integrators from commodity components. Almost all of these systems use Intel or AMD 32-bit microprocessor nodes and run Linux. The use of this type of cluster architecture was popularized by the Beowulf project,[21] following on previous Network of Workstations projects. Such Beowulf clusters are increasingly used as commercial capacity machines (e.g., Web servers and search engines) and as departmental or project scientific computing machines in research and industry. Such clusters are attractive because of their low purchase cost, the large number of component suppliers, and the ease of adding components. Clusters of this type, which use low-cost Ethernet interconnects, are often used to run "embarrassingly parallel" jobs consisting of many almost independent sequential subtasks.

The top U.S. vendors all offer clusters with 64-bit SMP nodes and custom switches. With the exception of HP, all provide custom switch interfaces. The top ranked Hewlett-Packard (HP) systems, including the second TOP500-ranked ASCI Q system, use AlphaServer SMP nodes connected (via a standard PCI interface) by a Quadrics switch; global communication uses message passing. Previous Hewlett-Packard clusters used the custom Hyperfabric interconnect. The top-ranked IBM systems, including the fourth-ranked (by TOP500) ASCI White system, use Power SMP nodes connected with an IBM proprietary switch using a proprietary interface (Power 4 systems currently use a standard I/O interface); global communication uses message passing. The Cray T3E uses Alpha uniprocessor nodes connected by a Cray proprietary switch with a proprietary interface that supports fast (put/get) remote memory access; the largest such system on the TOP500 list has 1,900 processors. (Cray is no longer pursuing the T3E architecture.) The SGI Origin uses MIPS quad-processor nodes connected with an SGI proprietary switch and an interface that supports cache-coherent global shared memory; the largest such system, with 1024 processors, is deployed at NASA Ames (SGI is now shipping systems that use Itanium processors). The Sun Fire, with up to 106 Sparc processors, also supports global cache-coherent shared memory.

NEC in Japan and Cray in the United States are at present the only vendors that manufacture vector processors for large-scale computing; their production volumes are significantly smaller than the volumes for nonvector 64-bit microprocessors. Such processors tend to be used in small-volume systems for the high end of the scientific and technical computing markets. In the past, other top Japanese vendors (Fujitsu and Hitachi) offered systems with vector processors. In the United States vector processors are being developed by Cray, with its new X1 product line. Ten systems on the current TOP500 list use Cray vector processor nodes.[22]

---

[21]Thomas Sterling, Donald J. Becker, Daniel Savarese, John E. Dorband, Udaya A. Ranawak, and Charles V. Packer. 1995. *Beowulf: A Parallel Workstation for Scientific Computation*. Proceedings of the 24th International Conference on Parallel Processing.

[22]Some mass-marketed microprocessors have limited support for vector instructions in a form that typically cannot be used to hide memory latency. The committee reserves the term "vector processor" for systems that have large vector register files and that support vector load/store instructions that can address noncontiguous memory locations.

The NEC Earth Simulator

The most significant Japanese supercomputer manufacturer is NEC. In the spring of 2002, NEC released the Earth Simulator (ES), a system with a peak performance of 40 Tflops/sec that is ranked first on the June 2003 TOP500 list. Based on the TOP500 LINPACK benchmark, the ES is the world's fastest computer by a factor of 2.58. An even greater ratio seems to hold for geosciences applications that it was specifically designed to support.

The ES is a cluster of 640 shared-memory multiprocessor (SMP) nodes. Each SMP node has eight processors, based on the SX-6 NEC processor design; each processor is a vector processor with a clock frequency of 500 MHz. Eight vector units within each processor provide a peak performance of 8 Gflop/sec per processor. The peak memory bandwidth is 32 GBps. Each processor has 72 vector registers, each with 256 elements. A robust crossbar network connects the nodes and provides a peak bandwidth of 16 GBps per node. The sustained bandwidth is approximately 12 GBps, full duplex. The design of the nodes of the ES (including vector processor and memory system) is evolutionary within the SX vector family. Semiconductor technology and advanced packaging are used to achieve performance. The software is also evolutionary and fairly stable.

It is instructive to compare the ES to the ASCI Q system at Los Alamos National Laboratory (LANL), which uses HP AlphaServer ES45 nodes and a Quadrics switch. Compared with the ASCI Q, the significant characteristics of the Earth Simulator are these:

- *Higher ratio of memory bandwidth to floating-point rate (4 B/flop versus 0.8 B/flop).* This ratio improves performance significantly for many codes that are memory intensive but do not exhibit the spatial and temporal locality exploited by caches. Although some such codes can be rewritten to be more cache-friendly, certain algorithms seem intrinsically difficult to localize.[23]
- *Use of vector parallelism in addition to SMP and message-passing parallelism.* The availability of a large number of vector registers and of vector load instructions makes it possible to prefetch data and to hide memory latency for codes where data accesses are predictable but not spatially localized. Codes that vectorize well can achieve a high fraction of the peak floating performance of the SX-6. On the other hand, the scalar performance of the SX-6 processor is not as good as the scalar performance of the Alpha processor, so the Alpha processor may achieve better performance on codes that do not vectorize well and are cache friendly.
- *Use of a global switch with a higher ratio of global bandwidth to floating-point rate (0.2 B/flop versus 0.03 B/flop).* This property contributes to performance on codes that require large amounts of global communication.

In summary, ES achieves a higher fraction of peak floating performance on many codes because of better memory bandwidth, better global bandwidth, and the availability of a memory prefetch mechanism (vector registers and vector load/store operations). There are no new micro-architectural concepts or unique technologies that are noteworthy in the ES. Rather, the performance is achieved through the use of a purpose-built microprocessor with high memory bandwidth and latency-hiding hardware and through the acceptance of a different budget balance between node hardware and interconnect hardware.

---

[23]See, for example, the GUPS benchmark described in Brian R. Gaeke, Parry Husbands, Xiaoye S Li, Leonid Oliker, Katherine A Yelick, and Rupak Biswas, 2002, *Memory-Intensive Benchmarks: IRAM vs. Cache-Based Machines*, International Parallel and Distributed Processing Symposium (IPDPS).

Software

Message passing is the main programming model used to scale applications to large systems; the MPI standard message-passing library is available on all TOP500 systems, including shared memory systems. Lower overhead communication can be achieved using put/get libraries on systems with suitable switch interfaces, such as the Cray T3E. Shared memory parallelism on SMP nodes is often exploited using OpenMP (i.e., C or FORTRAN with extensions for loop and task parallelism). However, OpenMP does not seem to be used for systemwide parallelism on large systems (even those supporting shared memory), perhaps because programmers lack the skill to use it well.

Almost all TOP500 systems use variants of Unix for their operating system. Shared memory systems are controlled by one global OS image, while distributed memory systems typically have one OS image per node. Lower-end Beowulf clusters typically use Linux, while higher-end systems use proprietary Unix systems. Libraries, programming tools, parallel file systems, and various system management tools complete the parallel programming environment available on these platforms.

Most vendor platforms use proprietary parallel programming environments. The proprietary software is often derived from open-source software; for example, all proprietary MPI implementations are derived from open source MPI implementations. Beowulf clusters mostly use open-source parallel software that is contributed by developers worldwide. Support for standard programming environments and interfaces, across all platforms, is an important goal that is only partially achieved. MPI and OpenMP are two successful standardization efforts in which industry adopted a de facto standard developed by the HPC community and/or the HPC vendors. Another successful model is provided by the TotalView parallel debugger, where a third-party software vendor supports the same software product across all main HPC platforms. However, TotalView is a singular example; attempts to standardize various tool interfaces and parallel system services, in particular parallel I/O, have had limited success. Although programming for large-scale parallel machines is more complex than programming for sequential machines, the typical programming environment available for scalable parallel computing is less sophisticated and less standardized than the environment available on small systems.

Algorithms

The algorithms used to run supercomputing applications are needed not just within the applications themselves but also to analyze the output data, store and transmit the data over unreliable media, load balance efficiently, and so on. The primary challenge introduced by supercomputing is that many conventional algorithms for these problems must be modified so as to scale effectively to much larger data sets or numbers of processors and to run efficiently on machines with deep memory hierarchies. For example, a numerical simulation on a very large mesh may involve converting an algorithm from one using dense matrices or even direct solvers on sparse matrices to one using a specialized iterative method that may still use a parallelized direct method on subproblems. Initially it may be possible to use a serialized mesh partitioner to load balance the matrix across processors, but as the matrix grows a parallel mesh partitioner may be needed. As another example, the problem may be scaled in order to introduce new physical models (e.g., one that respects polycrystalline structure in plasticity models), requiring wholly new discretizations and subgrid models. As this example illustrates, some of these algorithms are very specialized to particular application domains, whereas others, like mesh partitioners, are of quite general use.

3

Continuity and Predictability

The history of supercomputing in the U.S. in the last few decades is not a history of stability. Technical directions have changed (for instance, single instruction multiple data (SIMD) architectures have come and gone); tens of supercomputer manufacturing companies have gone out of business; models for government support of supercomputing R&D have changed; and levels of government support have fluctuated. Although technical fluctuations are caused, in part, by unpredictable changes due to innovation and may therefore be inevitable, more stability in supercomputing-related government policy would be advantageous.

A major policy issue for supercomputing, as well as for other technological fields, is the proper balance between investments that exploit and evolve current architectures and software and investments in alternative approaches that may lead to a paradigm shift. The interim report does not make recommendations on the crucial issue of how best to find that balance in the present context. This chapter and the next outline some of the main arguments for each type of investment. This chapter outlines the arguments for a continued, steady investment in the evolution of current architectures and software. The next chapter outlines the arguments for a sustained and vigorous research program in supercomputing. The two aspects are complementary.

IMPORTANT WORK IS GETTING DONE

In support of both basic scientific research and large missions of national importance, today's mix of supercomputing architectures and infrastructure are producing tangible results. For example, significant finding issues (SFIs) in the nuclear weapon stockpile have been closed with the help of massively parallel simulations run on ASCI platforms.[1]

Current missions require two kinds of supercomputer: (1) machines of high capability, on which a single very demanding problem uses the entire machine (for problems in which higher resolution in space or time is critical and for the most important and time-urgent problems, where faster times to solution are critical) and (2) workhorse-capacity machines that are used for multiple simultaneously executing jobs or embarrassingly parallel computations such as parameter studies. Capability machines stretch the scalability of current supercomputing technology to its limits—they are designed for the most demanding computational problems. Typically, the solutions to some of the computational problems within a mission organization will not exploit the full capability of such systems, either because the problem does

---

[1]As an example, an ASCI code has contributed to the yield reanalysis of a particular weapon that resulted in a revised certified yield.

not demand that level of capability or because current methods fail to achieve it. For those problems, systems using less aggressive technology and lower levels of parallelism will typically reach capacity at better overall cost and performance. However, the lesser systems will not provide the capability needed to solve critical problems in a timely manner.   Using ASCI codes and computers, the design of an arming, fusing, and firing device for the W76 warhead was optimized over a weekend. If run on serial machines, the same analysis would have taken significantly longer to complete.

NO NEAR-TERM ALTERNATIVES

One reason for fluctuations in investments in supercomputing is the perpetual hope for a silver bullet that will revolutionize supercomputing technology. While there are a number of promising opportunities for technology advances (e.g., processor in memory or streaming architectures), they are far in the future and will require concentrated investments to bring to fruition. They are also less than certain. Near-term research breakthroughs in high-performance computer architecture are unlikely, perhaps because of the lack of research investment in recent years. Evolution from current architectures is the only viable approach to meet needs in the immediate future (i.e., 3 to 5 years).

OLDER ARCHITECTURES COEXIST WITH NEW ONES

Changes do occur over time. For example, the "attack of the killer micros"[2] was largely successful, and supercomputers built of conventional microprocessors have largely displaced vector supercomputers. However, the current success of the Japanese Earth Simulator and the reentry of Cray in the vector market with the new X1 system also show the limitations of a short-term view: Supercomputers built of "killer micros" have not fully replaced vector architectures, even after many years, and the alternative technology still has its niche. Similarly, current supercomputing architectures are likely to be around for many years to come.

Furthermore, current architectures will survive and coexist with new innovations well after the latter are introduced. There has never been a single architecture that is best for all applications. Historically, the diverse needs of the U.S. scientific and defense missions have often led to the coexistence of major supercomputer architectures.

The evolution of supercomputer architecture probably will include supercomputers built from commercial microprocessor servers (such as the ASCI Q machine), hybrid systems that use commercial microprocessors with custom interfaces and switches (such as the T3E and the planned Red Storm system being built for Sandia National Laboratory), and purpose-built supercomputers (such as the Cray X1).

Building supercomputers from commodity components (as in the ASCI machines) will continue to be an attractive approach. The high-volume commodity market produces components, from processor chips to complete machines, with price-performance ratios difficult to achieve in any low-volume product. In the future, as in the past, there will continue to be opportunities to build larger machines with these components, whether as clusters of multiprocessors with standard I/O interfaces or as more integrated and higher bandwidth machines like Red Storm (using commodity AMD processors).

Scalable machines built from commodity parts have other attractive features in the context of a national supercomputing program. Because they are built on a scalable technology, the supercomputers are merely the extreme end of a continuum of products. Thus, some aspects of software and application

---

[2]Eugene Brooks. 1989. "Attack of the Killer Micros." White paper presented at Supercomputing 1989, Reno, Nev. A "killer micro" is a microprocessor-based machine that infringes on mini, mainframe, or supercomputer performance (see http://jargon.watson-net.com/jargon.asp?w=killer%20micro). The allusion is to the science fiction spoof *Attack of the Killer Tomatoes* (1978).

development can be done on smaller machines, affordable to an academic research group or department, for example, and then adapted to larger machines at national laboratories and centers.

Supercomputers built from custom processors also embody an evolutionary path with an implied human and technology investment over time. For example, the Cray X1 shares an architectural heritage with the Cray T3E, and its vector processors are a natural evolution from previous vector processor designs. Similarly, the NEC Earth Simulator should be seen not as a radically new design but as the natural evolution of earlier SX processors.

More customized supercomputers enlarge the set of applications that can be supported in a supercomputing universe. There is not a problem-independent, time-independent, or cost-independent dominance of custom-component supercomputers over commodity-component supercomputers, or vice versa. Each type contributes unique attributes to national-scale programs; the problem mix appropriate to each type can change as technology changes, but it is unlikely in the near future that either type will come to replace the other.

THE IMPORTANCE AND CONTINUING VALUE OF SOFTWARE RESEARCH AND ALGORITHM DEVELOPMENT

This committee notes, as many previous committees have noted, that software research and development continue to receive inadequate attention in national supercomputing programs. Hardware evolution needs to be supported by a robust investment in software development. Not only is such support necessary to ensure our ability to migrate key applications and algorithms to new architectures— indeed, if timely, it may inspire improvements in those architectures—but also it is necessary in support of the unique scaling requirements of supercomputing now and in the future. Government support for the development of the unique portable parallel debugger TotalView, which is marketed by Etnus and essential in supercomputing but of problematic economic viability in the broader computing marketplace, is a success story that needs to be repeated many times. Neither the current limited investments of platform vendors nor open source code developed by the national laboratories, industry, and academia are likely, by themselves, to fulfill the need for the standardized, high-quality programming environments that are needed to enhance programmer productivity in high-performance computing.

An interesting aspect of the Earth Simulator system is the successful use of high performance FORTRAN (HPF) on realistic applications to achieve significant performance levels (e.g., 14.9 Tflop on a three-dimensional fluid simulation for fusion science). HPF is an extension to FORTRAN that was developed in the early 1990s and was viewed at the time as a promising approach for achieving high performance on scalable computers while programming at a higher level of abstraction. However, HPF did not deliver (fast enough) on its early promises and was largely abandoned in the United States The successful use of HPF on the Earth Simulator suggests that there may have been a lack of perseverance in pursuing this and other software technologies. This is not to argue for the merits of HPF per se. Rather, it is to point out that lack of appreciation for the promise of supercomputing software technology may rob promising approaches of the time they need to mature.

A similar observation applies to application code development. Changing computing platforms require continual rethinking and redesigning of codes. Additionally, the continually increasing computing power produces new scientific targets of opportunity for which code enhancements must be made. Support for this activity has been consistently undervalued.

Investments in improvements to parallelism and in testing and validating algorithms and methods will continue to have value, even if supercomputing architectures change. For example, many software strategies for handling parallelism transcend the details of a particular parallel architecture. Additionally, over time, successors to some high-end machines tend to resemble widely available machines. Put most simply, tomorrow's Beowulf clusters may well look like today's ASCI machines. Hence, software research and development done on today's supercomputers will benefit tomorrow's smaller research

machines. Opportunities for leveraging people, talent, and experience across institutional boundaries can be exploited.

LEGACY CODES CANNOT BE ABANDONED UNTIL THEY ARE REPLACED

An important aspect of software evolution is the handling of existing scientific application codes (often referred to as legacy codes). Legacy codes represent a major investment; they evolve over many years, even decades, of use. Changes in programming languages, tools, and libraries, in programming models, or in hardware platforms entail a significant cost as codes are ported or rewritten. Several software capabilities are needed to support both the use and the evolution of legacy applications. Supercomputer system software must provide continuing support for the basic operations used in the applications by keeping the legacy software running until it can be replaced, by providing tools for performance tuning and debugging on new platforms, and by providing effective methods for porting and evolution.

Ideally, software should present a stable programming model. This allows programmers to be more productive in creating new software, because they can better leverage their experience. Moreover, developing new algorithms to exploit different costs of computation takes time. To the extent that a stable programming model reduces the need to reprogram fundamental computations, it can smooth this transition. It also mitigates the training costs.

Many changes in architecture have been accompanied by disruptions in programming models and software tools, impeding progress. It is imperative that new systems be developed in conjunction with development software (preferably using familiar, portable programming models) if they are to increase user productivity. Of particular importance is ensuring that changes in low-level hardware do not create unnecessary changes in higher-level software; for example, programmers working in object-oriented languages should not have to rewrite their code for new processor instruction sets. This will happen only by structuring funding to consider both hardware and software at the research stage and not allowing one to go forward without the other.

UNCERTAINTY AND INCONSISTENT POLICIES CAN BE EXPENSIVE

Failure to maintain steady, substantial investment in supercomputing could raise the cost of developing new generations of supercomputers in a timely manner. Rational firms will substantially reduce their level of effort when there is uncertainty about the demand for certain products. Because the cost of building up and tearing down the small, highly skilled teams that develop supercomputers can be significant, a temporary reduction in supercomputer acquisitions might in the long run raise the overall cost of supercomputing procurements.

It may be that the "keep the shipyard alive" argument—used to ensure that Navy ships are built at a continuous, predictable rate as a matter of national interest—applies to supercomputer acquisitions from multiple vendors.

While maintaining a predictable, continuous stream of investments in supercomputing is important, it appears that diversity of investment also is crucial (see Chapter 4). A rational investment program requires a diversified portfolio of investments in a variety of platforms, both hardware and software, and in basic research.

In summary, it is important that the long-term benefits of maintaining multiple suppliers, and helping them to do the long-range planning that sustains their supercomputing expertise, be carefully considered by policy makers.

4

Future Supercomputing and Research

As the uses of computing to address important societal problems continue to grow and the place of supercomputing within the overall computing industry continues to change, the value of innovation in supercomputer architecture, modeling, systems software, applications software, and algorithms will endure. Drawing on the recent supercomputing reports summarized in Chapter 2 and on its own insights, the committee outlines the main arguments for a vigorous research program in supercomputing. Some characteristics of successful innovation in areas such as high-performance computing are described and some key research problems are identified.

INNOVATION IN HIGH-END COMPUTING

A mature field is one characterized by small incremental improvements rather than large changes. By that measure, computing, and in particular high-end computing, is not a mature field. The underlying technology continues to evolve at a rapid pace,[1] and there are ample opportunities for innovations in architecture, systems software, and applications software to dramatically improve performance and productivity. New architecture and software technologies are needed to maintain historical growth in performance. To ensure that new technologies are available to form the basis for supercomputers in the 5-15 year time frame (a typical interval between research innovation and commercial deployment in the computer industry), a significant and continuous investment in basic research is required. Historically, such an investment in basic research has returned large dividends in terms of new technology. The need for basic research in supercomputing is particularly acute. Although there has been basic research in general-purpose computing technologies with broad markets, and there has been significant expenditure in advanced development efforts such as the ASC program and the TeraGrid, there has been relatively little investment in basic research in supercomputing architecture and software over the past decade, resulting in few innovations to be incorporated into today's supercomputer systems. Because supercomputing is affected by dislocations due to nonuniform technology scaling (described subsequently) before mainstream computers are affected, the lack of that research will eventually weigh on the broader computer industry as well.

Successful innovation programs in areas such as high-performance computing have a number of characteristics:

---

[1]International Technology Roadmap for Semiconductors. 2002. Update. Available online at <http://public.itrs.net/Files/2002Update/Home.pdf>.

22

- Continuous investment is needed at all stages of the technology pipeline, from initial investigation of new concepts to technology demonstrations and products. With no initial, speculative research, the pipeline dries out. With no technology demonstrations, new research ideas are much less likely to migrate to products. With no investment in products, supercomputers are not built.

- Continuous investment is needed in all contributing sectors, including universities, national laboratories, and vendors. All of these sectors, which often depend on government funding for their existence, have small communities of researchers and developers that are necessary for the continued evolution of supercomputing.

- A mix of small science projects and large efforts that create significant experimental prototypes is necessary. Large numbers of small individual projects are often the best way of studying new concepts. A smaller number of technology demonstration systems can draw on the successes of basic research in architecture, software, and applications concepts, demonstrate their interplay, and validate concepts ahead of their use in production systems. It is important that such pilot systems be built because without real hardware platforms, systems software and applications programs will not be written nor will the experience gained from such system building be acquired. For instance, pilot systems serve to identify research issues associated with the integration of hardware and software and to address system-level problems such as I/O performance in high-performance computing.

- Research is not a linear pipeline or a funnel, where losers are successively winnowed out until one winning product emerges. Successful research projects often incorporate the best ideas from related efforts. In addition, the experience gained from later stages often triggers reconsideration of earlier decisions. Good research should be organized to maximize the flow of ideas and people across projects and concepts.

ARCHITECTURE RESEARCH

The memory wall and the programming wall are two particularly challenging problems in supercomputing that could benefit from architecture research. The memory wall, the growing mismatch between memory bandwidth and latency and processor cycle time, is a major factor limiting performance. On many applications, modern processors are limited by memory system performance to a small fraction of peak performance.   This problem appears to be growing worse over time.

The memory wall is a special case of nonuniform scaling. As technology improves, different aspects of technology scale at different rates, leading to disparities in system parameters. When these disparities become large enough, or when a new technology is introduced, there is a discontinuity in system design that calls for innovative architecture and software.

To address the memory wall problem, innovative architectures are needed that increase memory bandwidth—or perhaps memory bandwidth per unit cost—both to local memory and across an interconnection network to remote memory. In addition, architectures must tolerate memory latency. Memory latency (local and global) is expected to increase relative to processor cycle time. Processors, however, can be designed to tolerate this latency without loss of performance by exploiting parallelism. While waiting for one result to return from memory, the processor works on a different, parallel part of the problem, perhaps using some combination of vectors and fine-grained multithreading. Innovation is needed to better identify and exploit locality. The memory bandwidth required by an application can often be reduced by a combination of architecture and software. The architecture provides local storage locations (registers and caches), and the software transforms the program to reduce the volume of intermediate data produced so that it fits in these local stores.

A programming wall also exists—namely, it is becoming increasingly difficult to write complex codes for high-end computers. Moreover, considerable effort is required to port these codes to new systems with different performance parameters. Innovation in architecture must take that issue into account. Architecture research is needed to devise a stable and efficient interface to systems software and applications, thereby masking, at least to some extent, the variety of architectural strategies that enhance

performance. For example, one could attempt to define an appropriate simple abstract virtual machine that reflected the architectural characteristics. All applications would then be programmed and compiled for that virtual machine. This approach could make it much easier for the software system to provide performance portability—that is, the ability to move a code from one machine to another without extensive performance tuning.[2] In addition, widespread adoption of better abstractions than those we have now would enable higher-level programming languages and simpler algorithm and software development.

Sequential orientation—the one-operation-at-a-time nature of contemporary architecture and software—is another major impediment to high-performance computing. While most high-end computers are necessarily parallel, they are built using processors and programming languages that are fundamentally serial. By incorporating notions of parallelism in the virtual machine described above, some issues of sequential orientation also might be mitigated.[3]

SOFTWARE RESEARCH

The development of scalable scientific codes today is a laborious process. Mathematical algorithms are translated by a programmer into detailed programs and tuned to a specific architecture using programming notations that reflect the underlying architecture—a manual, error-intensive process. The resulting code is hard to maintain, evolve, and port to new machines. The programmer must provide a wealth of detail that can obscure the high-level structure of the application solution—for example, the strategy to obtain parallelism. Also, the programmer may have an imperfect understanding of how low-level mechanisms are best used to achieve high performance.

High-performance computing offers unique challenges because of the need for large-scale parallelism and for latency tolerance and because performance is important for large, expensive hardware platforms. Research is needed to find fresh approaches to expressing both data and control parallelism at the application level, so that the strategy for achieving latency tolerance, locality, and parallelism is devised and expressed by the application developer, while separating out the low-level details that support particular platforms. Both new compilation and operating system capabilities and new tools are needed to realize high performance on modern supercomputing architectures.

Many of the languages, operating systems, and tools in current use have evolved by modifying languages and operating systems designed for sequential systems to infer opportunities for parallelism and to add explicit mechanisms to invoke parallel layouts and operations. Similarly, sequential tools have been modified for use in parallel environments. While that evolution is natural (and leverages existing knowledge and skills), it may no longer be sufficient. For example, the effort spent to maintain compatibility by changing the sequential base may limit the time available for enhancing support for parallelism and weaken the integrity of the parallel versions.

New software approaches for high-performance computing could exploit several special advantages. First, many application solutions are derived from a precise mathematical formulation of a physical problem, such as a finite difference discretization of a differential equation on a grid. By taking the problem domain into account, the necessary relationships between states of the computation and states of a mathematical system can facilitate both the mapping of the computation onto a large-scale parallel machine and the ensuing code development and testing. Second, HPC codes are often developed by small teams of highly capable scientists, who are often willing and able to use expert-friendly tools and environments if those tools will enhance their productivity. Finally, the difficulties in using the current tools for challenging and hard problems are a strong incentive for the user community to explore more advanced software technology.

---

[2]MPI is sometimes cited as such a target, but it is a low-level abstraction.

[3]Threads provide an aspect of concurrency, but possibly not in the form most appropriate for some applications developers and some hardware architectures.

It is important to pursue the most promising software research with perseverance and a long-term view. It often takes significant investments and a long time to change widely used programming paradigms. A significant investment also may be needed in compilers, run times, tools, and libraries to induce a user community to shift paradigms, even if the new paradigm holds the promise for significant productivity enhancements. An alternative approach, which is more common, is to fund a diversity of small enhancements to current systems. That approach runs the risk that none of the enhancements will make enough of a difference, and that few of them will be pushed far enough to be transferred to practice.

At the operating systems level, current HPC systems (especially cluster systems) have inherited a design that is not well tuned to the needs of large-scale parallel processing. As an example, a Linux cluster is managed by a large number of autonomous kernels, each making independent decisions on memory or processor allocation even though the entire cluster (or large parts of it) needs to run as one tightly coupled application. This discrepancy has been observed on many systems to have a negative effect on performance. A parallel application is not an entity that is recognized as a whole by the distributed operating system; there are no standardized parallel operating system services (e.g., parallel scheduling, parallel I/O, parallel memory management), although some implementations exist and are used. Communication and synchronization within a parallel application are achieved inefficiently using the same operating system mechanisms that are used for communication and synchronization across independent processes, or they bypass the system services. The mere need for bypass indicates that current interfaces are inadequate. Research that addresses the performance and semantic inadequacies of operating systems could lead to significant benefits in performance and software productivity and could push high-performance computing into new realms—for example, the use of large-scale parallelism for interactive computing.

Software research in HPC is likely to be more successful if closely coupled with research on algorithms and on architecture. Innovation often comes from a redesign of the interfaces between the various layers and from a better match between functionality across layers.

A major challenge in building revolutionary architectures and software systems is dealing with the large volume of legacy code. On the one hand, innovative research should not be constrained by compatibility needs of existing instruction sets, programming languages, operating systems, and application implementations. Advanced HPC research is likely to be more productive if it is free to explore paradigm-shifting approaches. On the other hand, a transition plan is needed to encourage adoption of new technologies. The transition plan should leverage the existing code base, because the cost of rewriting all of the legacy code from scratch is prohibitive.

RESEARCH ON APPLICATIONS AND ALGORITHMS

Supercomputing applications exist in a number of well-established and important fields, such as national security (cryptanalysis, intelligence, defense systems design, and nuclear stockpile stewardship), weather and climate forecasting, and automotive and aircraft design. New applications are emerging in the life sciences and biochemistry, among others.   This interim report was written in advance of an applications workshop to be held by the committee that will help it to formulate the supercomputing-related research needs in these areas. Following are some tentative observations.

There is an ever-increasing need for increased performance. The limitations of present-day supercomputers prevent many applications from being run using realistic parameter ranges of spatial resolution and time integration. For such applications, a significant increase of simulation and prediction quality can be attained by applying more computer power with primarily the same algorithms. For example, mesh resolution can be increased. In other applications, new algorithms and/or new processes are required to substantially advance the application involved. Increased mesh resolution often requires the development of new physics or algorithms for sub-grid-scale processes. In some cases, submodels of detailed processes may be required within a coarser mesh (e.g., cloud-resolving submodels embedded within a larger climate model grid).

As applications evolve, the workload characteristics change. Many codes evolve toward more complex, time-dependent and data-dependent control logic and more irregular data structures. This evolution taxes current architectures.

Huge amounts of model output and real data play an integral part of almost all supercomputing applications. The ways in which large datasets are prepared, stored, visualized, and analyzed highlight the need for new software, input/output, storage, and communication capabilities to go along with enhanced supercomputers and advanced methods.

Improvements in algorithms can sometimes improve performance much more than improvements in hardware and software do. For example, algorithms for solving the special linear system arising from the Poisson equation[4] on a regular grid have improved over time from needing $O(n^2)$ arithmetic operations to $O(n \log n)$ or even $O(n)$. Such algorithmic improvements can contribute to increased supercomputer performance as much as decades of hardware evolution. While such breakthroughs are hard to predict, the rewards can be significant. Further research can lead to such breakthroughs in the many complicated domains to which supercomputers are applied.

New algorithmic demands are driven by the following needs:

- *Disciplinary needs*. The need for higher-resolution analyses leads to larger problems to solve that lead to the need for faster algorithms (e.g., $O(n \log n)$ instead of $O(n^2)$). As the resolution increases, completely different physical models may be required (e.g., particle models instead of continuum models), which in turn require different solution methods. In some problems (such as turbulence), physically unresolved processes at small length or time scales may have large effects on macroscopic phenomena, requiring approximations that differ from those for the resolved processes.

- *Interdisciplinary needs*. Many real-world phenomena involve two or more coupled physical processes for which individual models and algorithms may be known (clouds, winds, ocean currents, heat flow inside and between the atmosphere and the ocean, atmospheric chemistry, and so on) but where the coupled system must be solved. Vastly differing time and length scales of the different disciplinary models frequently makes this coupled model much harder to solve.

- *Synthesis and optimization replacing analysis*. After one has a model that can be used to analyze (predict) the behavior of a physical system (such as an aircraft or weapons system), it is often desirable to use that model to try to synthesize or optimize a system so that it has certain desired properties. Such a problem can be much more challenging than analysis alone. As an example, a typical analysis computes, from the shape of an airplane wing, the lift resulting from air flow over the wing, by solving a differential equation. The related optimization problem is to choose the wing shape that maximizes lift, incorporating the constraints that ensure that the wing can be manufactured. Solving that problem requires determining the direction of change in wing shape that causes the lift to increase, either by repeating the analysis as changes to shape are tried or by analytically computing the appropriate change in shape.

- *Huge data sets*. Many fields (one is biology) that previously had relatively few quantitative data to analyze now have very large quantities, often of varying type, meaning, and uncertainty. These data may be represented by a diversity of data structures, including tables of numbers, irregular graphs, adaptive meshes, relational databases, two- or three-dimensional images, text, or various combined representations. Extracting scientific meaning from these data requires coupling numerical, statistical, and logical modeling techniques in ways that are unique to each discipline.

- *Changing machine models*. A machine model is the set of operations and their costs presented to the programmer by the underlying hardware and software. As the machine model changes between technology generations, an algorithm will probably have to be changed to maintain performance and scalability. This could involve adjusting a few parameters in the algorithm describing data layouts, running a combinatorial optimization scheme to rebalance the load, or using a completely different algorithm that trades off computation and communication in different ways. Some success has been

---

[4]A Poisson equation is an equation that describes many physical systems, including heat flow, fluid flow, diffusion, electrostatics, and gravity, with *n* unknowns.

achieved in automating this process, but only for a few important algorithmic kernels. For example, the ATLAS[5] and FFTW[6] systems automatically choose implementations of matrix-matrix-multiplication and the fast Fourier transform, respectively, to maximize performance on a particular architecture, depending on properties such as memory speed and number of registers.

Emerging application areas also drive the need for new algorithms and applications. Bioinformatics, for example, is driving the need to couple equation-driven numerical computing with probabilistic and constraint-driven computing. Large volumes of data from the Human Genome Project, clinical trials, statistics, population genetics, and imaging and visualization research stress the I/O capabilities of contemporary systems.

Many simulation and optimization codes that are now evolved and maintained by independent software vendors (ISVs) originated in research labs and universities. The arguments for government funding of supercomputing that are outlined in the next section apply as well to supercomputing application software: markets are likely to underinvest in such software, the government is an early and main customer for many such packages, and ISV codes are heavily used in weapon design. Thus, there is a strong case for government investment, not only in algorithm and application research, but also in the development of robust and scalable application software.

---

[5]See <http://math-atlas.sourceforge.net>.
[6]See <http://www.fftw.org>.

5

The Role of Government in Supercomputing

The federal government has been involved in the development and advancement of supercomputing since the advent of computers themselves. While the precise mechanism and level of support have varied over time, there has been a long-standing federal commitment to encourage the technical progress and diffusion of high-performance computing systems (some of this history is summarized in Chapter 2). Economic and policy analysis emphasizes several broad justifications for government involvement in technology development; the key justifications that apply to supercomputing are outlined below.

GOVERNMENT AS A LEADING CUSTOMER

Much technological innovation is (at least initially) directed toward applications dominated by government involvement and purchasing. Most notably, defense needs have often been the specific setting in which new technologies—including supercomputing—are first developed and applied. Even when commercial firms are purchasing or exploiting supercomputer technology, as in the pharmaceutical and biotechnology industries, governments are often the largest single customer for the resulting innovations (e.g., through the Medicare and Medicaid programs in the United States or national health insurance programs in other countries). The federal government remains the single largest purchaser of supercomputers in the world, for mission-oriented tasks ranging from national security to health to climate modeling.[1]

Some government missions may require specifications that are peculiar to the individual applications. In such cases, it may be important for the relevant federal agencies and research labs to be closely involved in the associated R&D (including prototyping), even when the research and development are carried out in the private sector.

In the United States, Japan, and elsewhere, the majority of supercomputers have been purchased directly or indirectly using government funds, and the committee has no evidence that this pattern is likely to change in the foreseeable future. As the social custodian of well-defined government missions and the largest and most aggressive customer for new technology related to these missions, the government has an incentive to ensure appropriate and effective funding for innovative investments in

---

[1]IDC estimates that the high-end HPC market in the United States has been around $1 billion (±$200 million) per year since 1994. The U.S. government (including DOD, the national laboratories, and classified programs) spends roughly $700 or $800 million per year (and this spending has been relatively flat for the last 10 years). SOURCE: Debra Goldfarb, IDC HPC industry analyst, presentation to the committee on May 23, 2003.

supercomputing technology so as to guarantee that the technology progresses at a rate and in a direction that serve government missions.

NATIONAL SECURITY IMPLICATIONS

U.S. government users require assured, secure, and reliable access to supercomputing capabilities to ensure that critical national security requirements are met. Supercomputers are used to develop intelligence gathering equipment (e.g., better antennas), to find important information through massive data mining, and for cryptography and cryptanalysis. They are used to design better weapons, airplanes, and tanks as well as for battlefield-related calculations that must be carried out very quickly. For example, the timely calculation of areas of enemy territory where enemy radars are not able to spot our airplanes (as was done during the first Gulf war) can be crucial. Design and refurbishment of nuclear weapons depends critically on supercomputing calculations, as does the design of next-generation armament for the Army's Future Combat System.

It is likely that supercomputing will be increasingly important to homeland security. Examples include micrometeorology analysis to combat biological terrorism and computer forensic analysis of terrorist bombings.

The federal government must be able to guarantee that such systems do what they are intended to do with no harmful side effects from, for instance, malicious insertion of incorrect calculations. It must guarantee that supercomputers are available to U.S. security agencies with no hindrance and must be able to restrict access to some supercomputing technologies abroad. All in all, the government has an incentive to ensure a strong supercomputing technology base in the United States.

MARKET FORCES

Economists are generally reluctant to see government intervene in highly competitive markets, where the costs of disruption to well-functioning and efficient resource allocation mechanisms are likely to be high. However, in many circumstances, market-based incentives for scientific discovery and innovation are likely to be insufficient. Because innovators often are unable to capture the full value of their inventions, market forces alone typically will bring less innovation than is worthwhile from society's perspective. Typically, underinvestment is greatest for basic research, fundamental scientific discoveries, or technologies that serve as stepping-stones for follow-on research by others.

A number of computing innovations first implemented in supercomputers (for example, instruction lookahead, multiple arithmetic units, multiple instruction buffers and data operators, pipelining, programmable I/O processors)[2] played an important role in shaping the architecture and performance of mainstream computers today (from workstations to personal computers). Initiatives funded in the context of supercomputers have influenced the ability to commercialize innovations, from workstation architecture to the latest Intel CPU. Machines equivalent in power to the supercomputer of 15 years ago that cost millions of dollars can now be purchased online for less than $1,000. The dramatic decrease in price for the same performance is due to continuing advances in semiconductor technology. However,

---

[2]For historical compilations of these and other major innovations in computer architecture, see Harold S. Stone et al., 1980, "Hardware Systems," in *What Can Be Automated? The Computer Science and Engineering Research Study,* Bruce W. Arden, ed., Cambridge, Mass.: MIT Press, pp. 319 and 320; C. Gordon Bell and Allen Newell, 1971, *Computer Structures: Readings and Examples,* New York, N.Y.: McGraw-Hill, fig. 2c, p. 45, p. 71; C. Gordon Bell and Allen Newell, 1982, *Computer Structures: Principles and Examples*, New York, N.Y.: McGraw-Hill, p. 393. These and other sources have been drawn together and annotated with additional information in K. Flamm, 1988, *Creating the Computer: Government, Industry, and High Technology*, Washington, D.C.: Brookings Institution Press. pp. 260-269.

the architecture of the chip and the software that drives it are based on the supercomputers of 15 years ago.

Supercomputing technology created with government funds has found its way into commercial computers and, in turn, has been used by private firms in basic research and to improve the design of products, including airplanes, automobiles, and pharmaceuticals. Supercomputing is essential for fundamental and applied research in materials science, earth science, life sciences, and so on. It is interesting to note that the Earth Simulator system that is described earlier is dedicated exclusively to the geosciences. The availability of such a powerful tool is likely to accelerate research in those disciplines by leading, for example, to a better understanding (and, hence, better prediction) of climate change and earthquakes. The value of such accelerated progress appears to be immense. It would seem, however, that only a small share of all the benefits is being captured by those involved in the initial stages of their development.

In summary, keeping the U.S. supercomputing industry at the technological cutting edge is vital for the security interests of the United States. In light of this imperative, there are several economic arguments that might justify continued government funding for supercomputing R&D rather than reliance on the marketplace alone.

6

Conclusion

The United States has always played a leadership role in bringing computing technology to bear on science and engineering research and development, advancing entirely new frontiers. The design of the earliest computers, for example—Britain's Colossus and the U.S.'s ENIAC during World War II[1]—was driven by defense and national security needs.   In the United States, ENIAC was followed by a steady stream of government-driven high-performance systems.[2] Supercomputing continues to be important for satisfying those needs.

The particular technical approaches of any program that develops or uses supercomputing represent a complex compromise between conflicting requirements and the risks and opportunities entailed in various approaches. As described earlier in this report, an assessment of the approaches requires a detailed understanding of (1) the applications, (2) the algorithms used to solve those applications, (3) codes and programming environments, (4) the performance of codes on various platforms, (5) the likely evolution of various hardware and software technologies under various funding scenarios, and (6) the costs, probabilities, and risks involved in various approaches.

In its final report, the committee will seek to characterize broadly the requirements of different application classes and to examine architecture, software, algorithm, and cost challenges and trade-offs associated with these application classes, keeping in mind the needs of the nuclear stockpile stewardship program, the broad science community, and the national security community. (Note that the identification of the distinct requirements of the stockpile stewardship program and its relation to the ASC acquisition strategy is expected to be the focus of a separate classified report by the JASONs). The committee believes it would be unwise to significantly redirect or reorient current supercomputing programs before careful scientific consideration has been given to the issues described above. Such changes might be hard to reverse, might reduce flexibility, and might increase costs in the future.

Exciting opportunities to advance knowledge and to serve society using supercomputing continue to emerge. The life and health sciences are becoming extraordinarily data rich, and researchers in those sciences are struggling to make sense of the data. The fruits of the genome projects for physiology and medicine cannot be realized without significant investments in computational hardware, algorithms, the

---

[1]Colossus was designed to decrypt German codes. See <http://www.codesandciphers.org.uk/lorenz/colossus.htm>. ENIAC (Electronic Numerical Integrator Analyzer and Computer) was built to calculate ballistic firing tables. See <http://ftp.arl.mil/~mike/comphist/eniac-story.html>.

[2]For a discussion of early government-funded projects in the late 1940s and 1950s that essentially created the early U.S. computer industry, see Chapter 3 ("Military Roots") in Kenneth Flamm, 1988, *Creating the Computer: Government, Industry, and High Technology*, Washington, D.C.: Brookings Institution Press.

software infrastructure, and the human infrastructure. The understanding of the physical world enabled by simulation and modeling is reaching ever-higher levels of fidelity and timeliness.

As in many other areas of technology R&D, there seem to be sound economic and social arguments for continued government investment in supercomputing. To sustain our leadership in supercomputing, to meet the security and defense needs of our nation, and to realize the opportunities to use supercomputing to advance knowledge, progress in supercomputing must go on. Continuity and stability in the government funding of supercomputing appear to be essential to the well-being of supercomputing in the United States.

An appropriate balance must be struck between evolutionary and innovative advances. Evolution is important because it allows present achievements to be exploited and because a diversity of approaches to supercomputing—including refinements of existing approaches—appears to be necessary to address the diversity of the computational challenges we face. Innovation in supercomputing stems from application-motivated research, which leads to experimentation and prototyping and then, in turn, to advanced development and testbeds and, finally, deployment and products. All the stages along that path need sustained investment. Coupled innovations in architecture, in software, in algorithms, and in application strategies and solution methods are equally important. Balance is also needed between exploiting cost-effective advances in widely used hardware and software and developing custom solutions that meet the most demanding needs. As we reach the limitations of current approaches and encounter the disruptions that are unavoidable when different technologies grow at different rates, the fruits of that research and its maturation into practice will prepare us for major paradigm shifts in the future.

Appendixes

A

Committee Member and Staff Biographies

COMMITTEE BIOGRAPHIES

Susan L. Graham (NAE), *Co-chair*, is the Pehong Chen Distinguished Professor of Electrical Engineering and Computer Science at the University of California, Berkeley, and the chief computer scientist of the National Partnership of Advanced Computations Infrastructure (NPACI). Her research spans many aspects of programming language implementation, software tools, software development environments, and high-performance computing. As a participant in the Berkeley Unix project, she and her students built the Berkeley Pascal system and the widely used program profiling tool gprof. Their paper on that tool was selected for the list of best papers from 20 years of the Conference on Programming Language Design and Implementation (1979-1999). She has done seminal research in compiler code generation and optimization. She and her students have built several interactive programming environments, yielding a variety of incremental analysis algorithms. Her current projects include the Titanium system for language and compiler support of explicitly parallel programs and the Harmonia framework for high-level interactive software development. Dr. Graham received an A.B. in mathematics from Harvard University and M.S. and Ph.D. degrees in computer science from Stanford University. She is a member of the National Academy of Engineering and a fellow of the Association for Computing Machinery (ACM), the American Association for the Advancement of Science (AAAS), and the American Academy of Arts and Sciences. In 2000 she received the ACM SIGPLAN Career Programming Language Achievement Award. In addition to teaching and research, she has been an active participant in the development of the computer science community, both nationally and internationally, over the past 25 years. She was the founding editor in chief of *ACM Transactions on Programming Languages and Systems*, which continued under her direction for 15 years. She has also served on the executive committee of the ACM special interest group on programming languages and as a member and chair of the ACM Turing Award committee. Dr. Graham has served on numerous national advisory committees, boards, and panels, including the National Research Council's (NRC's) Computer Science and Telecommunications Board, the NRC's Commission on Physical Sciences, Mathematics, and Applications, the advisory committee for the NSF Science and Technology Centers, and the advisory committee of the NSF Center for Molecular Biotechnology. Dr. Graham is a former member of the Presidential Information Technology Advisory Committee (PITAC).

Marc Snir, *Co-chair,* is Michael Faiman and Saburo Muroga Professor and head of the department of computer science at the University of Illinois at Urbana-Champaign. Dr. Snir's research interests include large-scale parallel and distributed systems, parallel computer architecture, and parallel programming. He received a Ph.D. in mathematics from the Hebrew University of Jerusalem in 1979 and worked at New York University (NYU) on the NYU Ultracomputer project from 1980 to 1982; at the Hebrew University of Jerusalem from 1982 to 1986; and at the IBM T.J. Watson Research Center from 1986 to 2001. At

IBM he headed research that led to the IBM scalable parallel system; contributed to Power 4 and Intel server architecture; and initiated the Blue Gene project. Dr. Snir has published more than a hundred papers on computational complexity, parallel algorithms, parallel architectures, interconnection networks, compilers, and parallel programming environments; he was a major contributor to the design of MPI. Dr. Snir is an ACM fellow and a fellow of the Institute of Electrical and Electronics Engineers (IEEE). He serves on the editorial boards of *Parallel Processing Letters* and *ACM Computing Surveys*.

William Dally received the B.S. degree in electrical engineering from Virginia Polytechnic Institute, the M.S. degree in electrical engineering from Stanford University, and the Ph.D. degree in computer science from Caltech. He is currently a professor of electrical engineering and computer science at Stanford University, where his group developed the Imagine processor, which introduced the concepts of stream processing and partitioned register organizations. Dr. Dally and his group have developed system architecture, network architecture, signaling, routing, and synchronization technology that can be found in most large parallel computers today. While at Bell Telephone Laboratories he contributed to the design of the BELLMAC32 microprocessor and designed the MARS hardware accelerator. At Caltech, he designed the MOSSIM Simulation Engine and the Torus Routing Chip, which pioneered wormhole routing and virtual-channel flow control. While a professor of electrical engineering and computer science at the Massachusetts Institute of Technology, his group built the J-Machine and the M-Machine, experimental parallel computer systems that pioneered the separation of mechanisms from programming models and demonstrated very low overhead synchronization and communication mechanisms. Dr. Dally has worked with Cray Research and Intel to incorporate many of these innovations in commercial parallel computers and with Avici Systems to incorporate this technology into Internet routers, and he cofounded Velio Communications to commercialize high-speed signaling technology. He is a fellow of the IEEE, a fellow of the ACM and has received numerous honors, including the ACM Maurice Wilkes award. He currently leads projects on high-speed signaling, computer architecture, and network architecture. He has published over 150 papers in these areas and is an author of the textbook *Digital Systems Engineering*.

James W. Demmel (NAE) joined the computer science division and mathematics department at the University of California, Berkeley, in 1990, where he holds a joint appointment as the Dr. Richard Carl Dehmel Distinguished Professor. He is also the chief scientist of the Center for Information Technology Research in the Interest of Society (CITRIS) and has worked to create the atmosphere of collaboration and communication that fosters an interdisciplinary approach to information technology research. Dr. Demmel is an expert on software and algorithms to facilitate computational science, having contributed to the software packages LAPACK, ScaLAPACK, BLAS, and SuperLU. He is an ACM fellow and an IEEE fellow and has been an invited speaker at the International Congress of Mathematicians. He received a B.S. in mathematics from Caltech in 1975 and a Ph.D. in computer science from the University of California, Berkeley, in 1983.

Jack J. Dongarra (NAE) is a University Distinguished Professor of Computer Science in the computer science department at the University of Tennessee, an adjunct R&D participant in the computer science and mathematics division at Oak Ridge National Laboratory (ORNL), and an adjunct professor in the computer science department at Rice University. He specializes in numerical algorithms in linear algebra, parallel computing, use of advanced computer architectures, programming methodology, and tools for parallel computers. His research includes the development, testing, and documentation of high-quality mathematical software. He has contributed to the design and implementation of the following open source software packages and systems: EISPACK, LINPACK, the BLAS, LAPACK, ScaLAPACK, Netlib, PVM, MPI, NetSolve, TOP500, ATLAS, and PAPI. He has published approximately 200 articles, papers, reports, and technical memoranda and is coauthor of several books. He is a fellow of the AAAS, the ACM, and the IEEE. He earned a B.S. in mathematics from Chicago State University in 1972. A year later he finished an M.S. in computer science from the Illinois Institute of Technology. He received

his Ph.D. in applied mathematics from the University of New Mexico in 1980. He worked at the Argonne National Laboratory until 1989, becoming a senior scientist.

Kenneth Flamm holds the Dean Rusk Chair in International Affairs at the University of Texas Lyndon B. Johnson (LBJ) School of International Affairs. He joined the LBJ School in 1998, is a 1973 honors graduate of Stanford University, and received a Ph.D. in economics from MIT in 1979. From 1993 to 1995, Dr. Flamm served as principal deputy assistant secretary of defense for economic security and as special assistant to the deputy secretary of defense for dual-use technology policy. Defense Secretary William J. Perry awarded him the department's distinguished public service medal in 1995. Prior to his service at the Defense Department, he spent 11 years as a senior fellow in the Foreign Policy Studies Program at the Brookings Institution. Dr. Flamm has been a professor of economics at the Instituto Tecnológico A. de México in Mexico City, the University of Massachusetts, and George Washington University. He has also been an adviser to the director general of income policy in the Mexican Ministry of Finance and a consultant to the Organization for Economic Cooperation and Development, the World Bank, the National Academy of Sciences, the Latin American Economic System, the U.S. Department of Defense, the U.S. Department of Justice, the U.S. Agency for International Development, and the Office of Technology Assessment of the U.S. Congress. Dr. Flamm, an expert on international trade and high technology industry, teaches classes in microeconomic theory, international trade, and defense economics.

Mary Jane Irwin (NAE) is Distinguished Professor of Computer Science and Engineering at the Pennsylvania State University. Her research and teaching interests include computer architecture, embedded and mobile computing systems design, low power design, and electronic design automation. Her research is supported by the National Science Foundation, the MARCO Gigascale Silicon Research Center, Intel Corporation, and Microsoft. She received an honorary doctorate from Chalmers University, Sweden, in 1997 and the Penn State Engineering Society's Premier Research Award in 2001. Dr. Irwin was named a fellow of the IEEE in 1995 and a fellow of the ACM in 1996. She is currently serving as chair of the National Science Foundation's Computer Information Sciences and Engineering Directorate's Advisory Committee, as a member of the Technical Advisory Board of the Army Research Laboratory, as the editor in chief of ACM's *Transactions on Design Automation of Electronic Systems,* and as an elected member of the Computing Research Association's board of directors. In the past she has served as an elected member of the IEEE Computer Society's board of governors, of ACM's council, and as vice president of the ACM. She served as general chair of the 1996 Federated Computing Research Conference, the 36th Design Automation Conference, and the 2002 International Symposium on Low Power Electronics and Design. Dr. Irwin received her M.S. (1975) and Ph.D. (1977) degrees in computer science from the University of Illinois, Urbana-Champaign.

Charles Koelbel is a research scientist in the computer science department at Rice University. Dr. Koelbel's area of expertise is in languages, compilers, and programming paradigms for parallel and distributed systems—in layman's terms, developing computer languages and algorithms that let several computers "talk" to each other and work together efficiently. He has contributed to many research projects while at Rice, mostly through the Center for Research on Parallel Computation, an NSF-funded Science and Technology Center with the mission to make parallel computation usable by scientists and engineers. These projects include the National Computational Science Alliance Technology Deployment Partners program, the Department of Defense's high-performance computing modernization program, and the FORTRAN D programming language project. He was executive director of the High Performance FORTRAN Forum, an effort to standardize a language for parallel computing. More recently, he served for 3 years as a program director at the National Science Foundation, where he was responsible for the Advanced Computational Research program and helped coordinate the Information Technology Research program. He is coauthor of *The High Performance Fortran Handbook*, MIT Press, 1993, and many papers and technical reports. He received his Ph.D. in computer science from Purdue University in 1990.

Butler W. Lampson (NAE) is an architect and a Distinguished Engineer at Microsoft Corporation and an adjunct professor of computer science and electrical engineering at the Massachusetts Institute of Technology. He was on the faculty at the University of California, Berkeley, at the computer science laboratory at Xerox PARC, and at Digital's Systems Research Center. Dr. Lampson has worked on computer architecture, local area networks, raster printers, page description languages, operating systems, remote procedure call, programming languages and their semantics, programming in large, fault-tolerant computing, transaction processing, computer security, and WYSIWYG editors. He was one of the designers of the SDS 940 time-sharing system, the Alto personal distributed computing system, the Xerox 9700 laser printer, two-phase commit protocols, the Autonet LAN, Microsoft Tablet PC software, and several programming languages. He received an A.B. from Harvard University, a Ph.D. in electrical engineering and computer science from the University of California at Berkeley, and honorary science doctorates from the Eidgenoessische Technische Hochschule, Zurich, and the University of Bologna. Dr. Lampson holds a number of patents on networks, security, raster printing, and transaction processing. He is a former member of the NRC's Computer Science and Telecommunications Board. He has served on numerous NRC committees, including the Committee on High Performance Computing and Communications: Status of a Major Initiative. He is a fellow of the Association for Computing Machinery and the American Academy of Arts and Sciences. He received ACM's Software Systems Award in 1984 for his work on the Alto, IEEE's Computer Pioneer award in 1996, the Turing Award in 1992, and the von Neumann Medal in 2001.

Robert F. Lucas is the director of the computational sciences division of the University of Southern California's Information Sciences Institute (ISI). He manages research in computer architecture, VLSI, compilers, and other software tools. Prior to joining ISI, he was the head of the High Performance Computing Research department in the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory. He oversaw work in scientific data management, visualization, numerical algorithms, and scientific applications. Prior to joining NERSC, Dr. Lucas was the deputy director of DARPA's Information Technology Office. He also served as DARPA's program manager for scalable computing systems and data-intensive computing. From 1988 to 1998, he was a member of the research staff of the Institute for Defense Analysis's Center for Computing Sciences. From 1979 to 1984, he was a member of the technical staff of the Hughes Aircraft Company. Dr. Lucas received B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University in 1980, 1983, and 1988 respectively.

Paul Messina retired in March 2002 from the California Institute of Technology (Caltech), where he was assistant vice president for scientific computing, director of Caltech's Center for Advanced Computing Research, and faculty associate in scientific computing. He also served as principal investigator for the Distributed Terascale facility and Extensible Terascale facility projects at Caltech and was coprincipal investigator of the National Virtual Observatory Project. Currently, Dr. Messina is a distinguished senior computer scientist (part time) at Argonne National Laboratory and a senior advisor on computing to the director general of CERN, in Geneva. During a leave from Caltech from January 1999 to December 2000, he was Director of the Office of Advanced Simulation and Computing for Defense Programs in the NNSA at DOE. In that capacity he had responsibility for managing the Accelerated Strategic Computing Initiative, the world's largest scientific computing program, which is defining the state of the art in that field. He holds the position of chief architect for the National Partnership for Advanced Computational Infrastructure (NPACI), a partnership established by the National Science Foundation and led by the University of California, San Diego. His recent interests focus on advanced computer architectures, especially their application to large-scale computations in science and engineering. He has also been active in high-speed networks, computer performance evaluation, and petaflop computing issues. Prior to his assignment at DOE, he led the computational and computer science component of Caltech's research project, funded by the Academic Strategic Alliances Program (ASAP) of the ASCI (now called the Advanced Simulation and Computing Program). In the mid-1990s he established and led the Scalable I/O

Initiative (SIO). In the early 1990s, he was the principal investigator and project manager of the CASA gigabit network testbed. During that period he also conceived, formed, and led the Consortium for Concurrent Supercomputing, whose 13 members included several federal agencies, the national laboratories, universities, and industry. That consortium created and operated the Intel Touchstone Delta System, which was the world's most powerful scientific computer for 2 years. He also held a joint appointment at the Jet Propulsion Laboratory as manager of high-performance computing and communications from 1988 to 1998. Dr. Messina received a Ph.D. in mathematics in 1972 and an M.S. in applied mathematics in 1967, both from the University of Cincinnati, and a B.A. in mathematics in 1965 from the College of Wooster. He is a member of the IEEE Computer Society, the AAAS, the ACM, the Society for Industrial and Applied Mathematics, and Sigma Xi. He is coauthor of four books on scientific computing and editor of more than a dozen others.

Jeffrey M. Perloff is a professor of agricultural and resource economics at the University of California at Berkeley. His economics research covers industrial organization and antitrust, labor, trade, and econometrics. His textbooks are *Modern Industrial Organization* (coauthored with Dennis Carlton) and *Microeconomics*. He has been an editor of *Industrial Relations* and associate editor of the *American Journal of Agricultural Economics* and is an associate editor of the *Journal of Productivity Analysis*. He has consulted with nonprofit organizations and government agencies (including the Federal Trade Commission and the Departments of Commerce, Justice, and Agriculture) on topics ranging from a case of alleged Japanese television dumping to the evaluation of social programs. He has also conducted research in psychology. Dr. Perloff is a fellow of the American Agricultural Economics Association. He received his B.A. in economics from the University of Chicago in 1972 and his Ph.D. in economics from the Massachusetts Institute of Technology in 1976. He was previously an assistant professor in the Department of Economics at the University of Pennsylvania.

William H. Press (NAS) is deputy laboratory director for science and technology at the Los Alamos National Laboratory (LANL). Before joining LANL in 1998, he was professor of astronomy and physics at Harvard University and a member of the theoretical astrophysics group of the Harvard-Smithsonian Center for Astrophysics. He is also the coauthor and comaintainer of the *Numerical Recipes* series of books on scientific computer programming. Dr. Press was assistant professor of physics at Princeton University and Richard Chace Tolman Research fellow in theoretical physics at Caltech, where he received a Ph.D. in physics in 1972. He is a member of the National Academy of Sciences and was a founding member of its Computer and Information Sciences Section. He has published more than 140 papers in the areas of theoretical astrophysics, cosmology, and computational algorithms. He is also a fellow in the American Academy of Arts and Sciences, a member of the Council on Foreign Relations, and a past recipient of an Alfred P. Sloan Foundation fellowship and the Helen B. Warner Prize of the American Astronomical Society. Dr. Press is a past co-chair of the Commission on Physical Sciences, Mathematics, and Applications (CPSMA) of the National Research Council (NRC); a past member of the Chief of Naval Operations' Executive Panel, the U.S. Defense Science Board, the NRC's Computer Science and Telecommunications Board, the Astronomy and Astrophysics Survey Committee, and a variety of other boards and committees. He has led national studies in subjects including high-bandwidth telecommunications (the Global Grid), national science and technology centers (especially for computational science), and a wide variety of national security issues. Dr. Press serves as a scientific advisor to the David and Lucille Packard Foundation and other foundations. He is a member of the board of trustees of the Institute for Defense Analyses (IDA) and serves on its executive committee and on the external advisory committees of its CCS and CCR Divisions. He serves on the Defense Threat Reduction Agency's Science and Technology Panel.

Albert Semtner is a professor of oceanography at the Naval Postgraduate School in Monterey, California. He received a B.S. in mathematics from Caltech and a Ph.D. in geophysical fluid dynamics from Princeton. His prior professional positions were in UCLA's Meteorology Department and in the

Climate Change Research Section of the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. His interests are in global ocean and climate modeling and in supercomputing. Dr. Semtner has written extensive oceanographic codes in assembly language for shipboard use. He produced the first vectorized (FORTRAN) version of a standard ocean model in 1974 and the first parallel-vector version (in collaboration with Robert Chervin of NCAR) in 1987. He interacted with Los Alamos scientists on transitioning the parallel-vector code to massively parallel architectures in the early 1990s. Under the leadership of Warren Washington of NCAR, he participated in the development of the DOE Parallel Climate Model using the Los Alamos Parallel Ocean Program and a parallel sea ice model from the Naval Postgraduate School. That climate model has been ported to numerous parallel architectures and used as a workhorse climate model in numerous scientific applications. Dr. Semtner has been an affiliate scientist with NCAR for the last 12 years and simultaneously a member (and usually chair) of the Advisory Panel to the NCAR Scientific Computing Division. He is a winner (with R. Chervin) of a 1990 Gigaflop Achievement Award (for the vector-parallel code) and the 1993 Computerworld-Smithsonian Leadership Award in Breakthrough Computational Science (for global ocean modeling studies that included ocean eddies for the first time). Dr. Semtner is an associate editor of *Ocean Modeling* and of the *Journal of Climate*. He is also a fellow of the American Meteorological Society.

Scott Stern graduated with a B.A. degree in economics from New York University. After working for a consulting company in New York, he attended Stanford University and received his Ph.D. in economics in 1996. From 1995 to 2001, Dr. Stern was assistant professor of management at the Sloan School at MIT. Since September 2001, he has been an associate professor in the Kellogg School of Management at Northwestern University, a nonresident senior fellow of the Brookings Institution, and a faculty research fellow of the National Bureau of Economic Research. He is also a co-organizer of the Innovation Policy and the Economy Program at the National Bureau of Economic Research and an associate editor of *Management Science*. Dr. Stern explores how innovation—the production and distribution of ideas—differs from the production and distribution of more traditional economic goods and the implications of these differences for both business and public policy. Often focusing on the pharmaceutical and biotechnology industries, this research is at the intersection of industrial organization and economics of technological innovation. Specifically, recent studies examine the determinants of R&D productivity, the impact of incentives on R&D organization, the mechanisms by which firms earn economic returns from innovation, and the consequences of technological innovation on product market competition. A key conclusion from this research is that translating ideas into competitive advantage requires a distinct and nuanced set of resources and strategies. Effective management of innovation therefore requires careful attention to the firm's internal ability to develop truly distinct technologies and to subtle elements of the firm's external development and commercialization environment.

Shankar Subramaniam is a professor of bioengineering, chemistry, and biochemistry and biology and director of the Bioinformatics Graduate Program at the University of California at San Diego. He also holds adjunct professorships at the Salk Institute for biological studies and the San Diego Supercomputer Center. Prior to moving to the University of California, San Diego, Dr. Subramaniam was a professor of biophysics, biochemistry, molecular and integrative physiology, chemical engineering, and electrical and computer engineering at the University of Illinois at Urbana-Champaign (UIUC). He was also the director of the Bioinformatics and Computational Biology Program at the National Center for Supercomputing Applications and co-director of the W.M. Keck Center for Comparative and Functional Genomics at UIUC. He is a fellow of the American Institute for Medical and Biological Engineering (AIMBE) and is a recipient of Smithsonian Foundation and Association of Laboratory Automation Awards. Dr. Subramaniam has played a key role in raising national awareness of training and research in bioinformatics. He served as a member of the National Institutes of Health (NIH) Director's Advisory Committee on Bioinformatics, which produced the report *Biomedical Information Science and Technology Initiative* (BISTI). The report recognized the dire need for trained professionals in bioinformatics and recommended the launching of a strong NIH funding initiative. Dr. Subramaniam

serves as the chair of a NIH BISTI Study Section. Dr. Subramaniam has also served on bioinformatics and biotechnology advisory councils for Virginia Tech, the University of Illinois at Chicago, and on the scientific advisory board of several biotech and bioinformatics companies. Dr. Subramaniam served as review panel member of NIH CIT and his focus was on how CIT should respond to the BISTI initiative. Dr. Subramaniam has served as a member of the state of Illinois Governor's initiative in biotechnology and as advisor and reviewer of the state of North Carolina initiative in biotechnology. Dr. Subramaniam has published more than a hundred papers in the interdisciplinary areas of chemistry/biophysics/biochemistry/bioinformatics and computer science.

Lawrence C. Tarbell, Jr., is the deputy director of the Technology Futures Office for Eagle Alliance, a company formed in 2001 by the Computer Sciences Corporation and Northrop Grumman to outsource part of the IT infrastructure (workstations, local area networks, servers, and telephony) for the National Security Agency (NSA). His particular area of responsibility is IT enterprise management, with backup responsibility in distributed computing and storage. Mr. Tarbell spent the previous 35 years at NSA with responsibilities for research and development of high-performance workstations, networks, computer security, mass storage systems, and systems software. For over 13 years, he managed and led supercomputing research and applications development for NSA, sponsoring high-performance computing and mass storage research (both independently and jointly with DARPA and NASA) at many U.S. companies and universities. In 1990, he co-chaired Frontiers of Supercomputing II, sponsored jointly by NSA and Los Alamos National Laboratory. Mr. Tarbell received his M.S. in electrical engineering from the University of Maryland and his B.S. in electrical engineering (magna cum laude) from Louisiana State University.

Steven J. Wallach (NAE) is vice president of technology for Chiaro Networks, an advisor to CenterPoint Venture Partners, and a consultant to the U.S. Department of Energy ASC program. Chiaro Networks provides major disruptive technologies in a high-end routing platform for reliability, scalability, and flexibility. Previously, he was cofounder, chief technology officer, and senior vice president of development of Convex Computers. After Hewlett-Packard bought Convex, Mr. Wallach became the chief technology officer of HP's Large Systems Group. He was a visiting professor at Rice University from 1998 to 1999 and manager of advanced development at Data General from 1975 to 1981. He was the principal architect of the 32-bit Eclipse MV supercomputer and, as part of this effort, participated in the design of the MV/6000, MV/8000, and MV/10000 (chronicled in the Pulitzer Prize-winning book *The Soul of a New Machine*, by Tracy Kidder). Mr. Wallach was an engineer at Raytheon from 1971 to 1975, where he participated in various hardware design efforts, including the computer used to control the launching of the Patriot missile system and various signal processors. He had primary responsibility for the design of the all applications digital computer (AADC), which was intended for military specification airborne applications and was made up of gate arrays (one of the first such systems) and a vector instruction set based on APL. Mr. Wallach holds 33 patents. He was a member of the President's Information Technology Advisory Committee (PITAC). He holds a B.S. in engineering from the Polytechnic Institute of Brooklyn, an M.S.E.E. from the University of Pennsylvania, and an M.B.A. from Boston University.

STAFF BIOGRAPHIES

Cynthia A. Patterson is a study director and program officer with the Computer Science and Telecommunications Board of the National Academies. She is currently involved in a diverse set of CSTB projects, including a project on critical information infrastructure protection and the law, a study on the future of supercomputing, and a study on telecommunications research and development. She just completed a project that outlined a research agenda at the intersection of geospatial information and computer science and a joint study with the Board on Earth Sciences and Resources (BESR) and the

Board on Atmospheric Sciences and Climate (BASC) on public-private partnerships in the provision of weather and climate services. She has also been involved with the congressionally mandated study on Internet searching and the domain name system. Prior to joining CSTB, Ms. Patterson completed an M.Sc. from the Sam Nunn School of International Affairs at the Georgia Institute of Technology. Her graduate work was supported by the Department of Defense and SAIC. In a previous life, Ms. Patterson was employed by IBM as an IT consultant for both federal government and private industry clients. Her work included application development, database administration, network administration, and project management. She received a B.Sc. in computer science from the University of Missouri-Rolla.

Phil Hilliard is a research associate with the Computer Science and Telecommunications Board. He provides research support as part of the professional staff and is working on projects focusing on telecommunications research, supercomputing, and dependable systems. Before joining the National Academies, Mr. Hilliard worked at BellSouth in Atlanta, Georgia, as a competitive intelligence analyst and at NCR as a technical writer and trainer. He earned an M.B.A. from Georgia State University (2000) in Atlanta, Georgia, and a B.S. in computer and information technology from Georgia Institute of Technology (1986) in Atlanta, Georgia. He is currently working on his master's of library and information science in Florida State University's online program.

Margaret Huynh, senior project assistant, has been with CSTB since January 1999, working on several projects. She is currently working on the future of supercomputing, wireless technology prospects and policy, and Internet navigation and the domain name system and worked on the report *Beyond Productivity: Information Technology*, *Innovation, and Creativity*. She previously worked on the projects that produced the reports *IT Roadmap to a Geospatial Future, Building a Workforce for the Information Economy*, and *The Digital Dilemma: Intellectual Property in the Information Age*. Ms. Huynh assisted on the project Exploring Information Technology Issues for the Behavioral and Social Sciences (Digital Divide and Democracy). Ms. Huynh assists on other projects as needed. Prior to coming to the National Academies, Ms. Huynh worked as a meeting assistant at Management for Meetings for 4 months and as a meeting assistant at the American Society for Civil Engineers from September 1996 to April 1998. Ms. Huynh has a B.A. (1990) in liberal studies with minors in sociology and psychology from Salisbury University, Salisbury, Maryland.

B

Acronyms

ACP	advanced cyberinfrastructure program
AMD	Advanced micro devices
ASC	Advanced Simulation and Computing [formerly Accelerated Strategic Computing Initiative (ASCI)]
ASCI	Accelerated Strategic Computing Initiative
COTS	commercial off-the-shelf
CSTB	Computer Science and Telecommunications Board
DARPA	Defense Advanced Research Projects Agency
DDR&E	Director of Defense Research and Engineering
DOD	Department of Defense
DOE	Department of Energy
ES	Earth Simulator
GAO	Government Accounting Office
HPC	high-performance computer
HPCCI	High Performance Computing and Communications Initiative
HPCS	high productivity computing systems
IHEC	integrated high-end computing
ISV	independent software vendor
ITRD	Information Technology Research and Development
JASONs	a group of scientific advisors that provides the federal government with largely classified analyses
LANL	Los Alamos National Laboratory
LINPACK	library of FORTRAN 77 routines for solving problems in numerical linear algebra. It is used for benchmarking the speed of supercomputers.
MPI	message passing interface
NIH	National Institutes of Health
NNSA	National Nuclear Security Administration

43

NSA	National Security Agency
NSB	National Science Board
NSF	National Science Foundation
PACI	Partnerships for Advanced Computational Infrastructure
PCI	peripheral component interconnect or interface
PITAC	President's Information Technology Advisory Committee
SFI	Significant Finding Issue
SIMD	single instruction multiple data
SMP	shared-memory multiprocessor

C

Briefers to the Committee

March 6-7, 2003
Washington, D.C.

George Cotter, National Security Agency
John Crawford, Intel
Robert Graybill, DARPA
John Grosh, Office of the Secretary of Defense
Daniel Hitchcock, DOE, Office of Advanced Scientific Computing Research (OS)
Gary Hughes, National Security Agency
David Kahaner, Asian Technology Information Program
Jacob V. Maizel, Jr., National Cancer Institute
José Muñoz, DOE, Office of Advanced Simulation and Computing (ASCI)
Clay Sell, Clerk of the Senate Subcommittee on Energy and Water Development
David Turek, IBM

May 21-23, 2003
Stanford, California

Greg Astfalk, Hewlett-Packard
Gordon Bell, Microsoft Research
Debra Goldfarb, IDC
James Gray, Microsoft Research
John Levesque, Cray Inc.
John Lewis, Boeing
Scott McClellan, Hewlett-Packard
William Reed, DOE, Office of Advanced Simulation and Computing (ASCI)
Mark Seager, Lawrence Livermore National Laboratory
Burton Smith, Cray Inc.

45

What Is CSTB?

As a part of the National Research Council, the Computer Science and Telecommunications Board (CSTB) was established in 1986 to provide independent advice to the federal government on technical and public policy issues relating to computing and communications. Composed of leaders from industry and academia, CSTB conducts studies of critical national issues and makes recommendations to government, industry, and academia. CSTB also provides a neutral meeting ground for consideration of complex issues where resolution and action may be premature. It convenes discussions that bring together principals from the public and private sectors, assuring consideration of key perspectives. The majority of CSTB's work is requested by federal agencies and Congress, consistent with its National Academies context.

A pioneer in framing and analyzing Internet policy issues, CSTB is unique in its comprehensive scope and its effective, interdisciplinary appraisal of technical, economic, social, and policy issues. Beginning with early work in computer and communications security, cyber-assurance and information systems trustworthiness have been a cross-cutting theme in CSTB's work. CSTB has produced several reports known as classics in the field, and it continues to address these topics as they grow in importance.

To do its work, CSTB draws on some of the best minds in the country and from around the world, inviting experts to participate in its projects as a public service. Studies are conducted by balanced committees without direct financial interests in the topics they are addressing. Those committees meet, confer electronically, and build analyses through their deliberations. Additional expertise is tapped in a rigorous process of review and critique, further enhancing the quality of CSTB reports. By engaging groups of principals, CSTB gets the facts and insights critical to assessing key issues.

The mission of CSTB is to

- *Respond to requests* from the government, nonprofit organizations, and private industry for advice on computer and telecommunications issues and from the government for advice on computer and telecommunications systems planning, utilization, and modernization;
- *Monitor and promote the health of the fields* of computer science and telecommunications, with attention to issues of human resources, information infrastructure, and societal impacts;
- *Initiate and conduct studies* involving computer science, technology, and telecommunications as critical resources; and
- *Foster interaction* among the disciplines underlying computing and telecommunications technologies and other fields, at large and within the National Academies.

CSTB projects address a diverse range of topics affected by the evolution of information technology. Recently completed reports include *Cybersecurity Today and Tomorrow: Pay Now or Pay Later*; *Youth, Pornography, and the Internet*; *Broadband: Bringing Home the Bits*; *The Digital Dilemma: Intellectual Property in the Information Age*; *IDs—Not That Easy: Questions About Nationwide Identity Systems*; *The Internet Under Crisis Conditions: Learning from September 11*; and *IT Roadmap to a Geospatial Future*. For further information about CSTB reports and active projects, see <http://cstb.org>.