

Free Executive Summary

The Potential Impact of High-End Capability Computing on Four Illustrative Fields of Science and Engineering

Committee on the Potential Impact of High-End Computing on Illustrative Fields of Science and Engineering, National Research Council

ISBN: 978-0-309-12485-0, 154 pages, 8 1/2 x 11, paperback (2008)



This free executive summary is provided by the National Academies as part of our mission to educate the world on issues of science, engineering, and health. If you are interested in reading the full book, please visit us online at <http://www.nap.edu/catalog/12451.html>. You may browse and search the full, authoritative version for free; you may also purchase a print or electronic version of the book. If you have questions or just want more information about the books published by the National Academies Press, please contact our customer service department toll-free at 888-624-8373.

Many federal funding requests for more advanced computer resources assume implicitly that greater computing power creates opportunities for advancement in science and engineering. This has often been a good assumption. Given stringent pressures on the federal budget, the White House Office of Management and Budget (OMB) and Office of Science and Technology Policy (OSTP) are seeking an improved approach to the formulation and review of requests from the agencies for new computing funds. The study considered, as examples, four fields of science and engineering to determine which of their major challenges are critically dependent on high-end capability computing (HECC): atmospheric science, astrophysics, chemical separation, and evolutionary biology. While this study does identify the potential impact of HECC in these four fields, and thus implicitly identifies some potential funding opportunities, that is not the goal, and this study is no substitute for competitive review of specific proposals. Rather, the study is meant to illustrate the sort of examination that any field or federal agency could undertake in order to analyze the HECC infrastructure it needs to support progress toward its research goals, within the context of other means of attacking those goals.

This executive summary plus thousands more available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved. Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

Summary

INTRODUCTION

Many federal funding requests for more advanced computer resources assume implicitly that greater computing power creates opportunities for advancement in science and engineering. This has often been a good assumption. Given stringent pressures on the federal budget, the White House Office of Management and Budget (OMB) and its Office of Science and Technology Policy (OSTP) are seeking an improved approach to the formulation and review of requests from the agencies for new computing funds. The study that produced this report was commissioned by the Networking and Information Technology Research and Development (NITRD) program, which operates under the OSTP to coordinate federal investments in networking and information technology. The study addressed the charge shown in the Preface.

The study considered, as examples, four fields of science and engineering to determine which of their major challenges are critically dependent on high-end capability computing (HECC). The fields chosen for the study were the atmospheric sciences, astrophysics, chemical separations, and evolutionary biology. The committee found continuing demands from the four fields for more, and more powerful, high-end computing. All four areas rely on HECC to carry out simulations of systems that are too complex to analyze through observation, experiment, or theory. Three of the four areas (the exception being chemical separations) are dealing with very large amounts of data and need HECC to handle them.

“High-end capability computing” means advanced computing that pushes the bounds of what is computationally feasible. While that is often interpreted in terms of raw processing power, users expect to achieve new scientific understanding or engineering capabilities through computation. Processing power is just one means to that end. Computational science and engineering is a systems process, bringing together hardware, software, investigators, data, and other components of infrastructure in order to gain insight into some question. Thus, from the user’s perspective, high-end capability computing means whatever sort of advanced, nonroutine computing system is needed to push the computational science or engineering capabilities of a given field. It will always entail more risk and require more innovation than commodity computing, but not necessarily a novel computing platform. This report uses the term

“high-end capability computing,” or HECC, as shorthand for this nonroutine frontier of computation, the precise definition of which will vary by field.

While this study does identify the potential impact of HECC in these four fields, and thus implicitly identifies some potential funding opportunities, that is not the goal, and this study is no substitute for competitive review of specific proposals. Rather, the study is meant to illustrate the sort of examination that any field or federal agency could undertake in order to analyze the HECC infrastructure it needs to support progress toward its research goals, within the context of other means of pursuing those goals.

SUMMARY OF THE MAJOR CHALLENGES IN THE FOUR FIELDS

Astrophysics

A small sample of some of the most important discoveries in astrophysics made in the past decade includes dark matter and dark energy, exosolar planets, and good evidence for the existence of black holes. These discoveries have positioned the field to address the following major challenges:

1. What is dark matter?
2. What is the nature of dark energy?
3. How did galaxies, quasars, and supermassive black holes form from the initial conditions in the early Universe observed by the Wilkinson Microwave Anisotropy Probe (WMAP) and the Cosmic Background Explorer (COBE), and how have they evolved since then?
4. How do stars and planets form, and how do they evolve?
5. What is the mechanism for supernovae and gamma-ray bursts, the most energetic events in the known Universe?
6. Can we predict what the Universe will look like when observed in gravitational waves?

To answer the questions posed by Challenges 3-6, advances in HECC are necessary. Challenges 1-2 are limited in the near term by the need for advanced astronomical observations, but these observations will produce so much data that HECC will in any case be needed for their analysis. The current situation for these challenges is described in Chapter 2.

While astrophysics is a computationally mature discipline—that is, it has a long history in the use of computing to solve problems—it would certainly benefit from access to more, and more powerful, HECC resources. The primary computational challenge is associated with the enormous dynamic range in length scales and timescales needed to resolve astrophysical processes. For example, grids as large as 2048^3 are currently used for calculations involving hydrodynamics, but even then the spatial features they are able to resolve are only about a hundredth the size of the computational domain. The availability of systems of 10^5 or more processors will enable much larger calculations (encompassing finer resolution, models of more physical processes, or both) while also making it feasible to perform more complex calculations that couple different models. The community needs support for porting its codes to multicore and petascale environments.

The committee identified some likely ramifications of inadequate or delayed support of HECC for astrophysics:

- The rate of new discovery would be limited.
- Inadequate support for HECC would lead to a failure to optimize investment in expensive experimental and observational facilities.

- Data are likely to be underexploited. Without enough HECC, the data collected by large-scale surveys cannot be properly managed and analyzed, so their full potential cannot be realized.

The Atmospheric Sciences

Weather forecasting and climate simulation require detailed simulations of the atmosphere. The skill and reliability of forecasts has increased markedly since the advent of weather radar, earth-observing satellites, and powerful computers. To look beyond a few hours, we combine powerful computers with the relevant laws of physics converted into mathematical models to predict how the observed present state of the global atmosphere will evolve in the hours and days ahead. We expect that major improvement will be achieved with much higher resolution and sophistication in the numerical models that portray events in the atmosphere and ocean and on the land surface. Numerical weather forecasting, in particular, will soon be able to take account of local features (e.g., lakes, ridges) and local variations in atmospheric moisture content, and the successful modeling of atmospheric variables on these scales should lead to a leap forward in forecast quality. Climate simulation and prediction requires detailed treatments of the physics, chemistry, and biology of the atmosphere, ocean, and land surface. Feedbacks in the integrated Earth system require high spatial resolution, as is the case for weather forecasting, but also additional mechanistic models of chemical and biological interactions.

The major challenges facing the atmospheric sciences are these:

1. Extend the range, accuracy, and utility of weather prediction. [1]
2. Improve our understanding and the timely prediction of severe weather, pollution, and climate events. [1]
3. Improve understanding and prediction of seasonal, decadal, and century-scale climate variation on global, regional, and local scales. [1]
4. Understand the physics and dynamics of clouds, aerosols, and precipitation. [2]
5. Understand the atmospheric forcing and feedbacks associated with moisture and chemical exchange at Earth's surface. [1]
6. Develop a theoretical understanding of nonlinear bifurcation and tipping points in weather and climate systems. [2]
7. Create the ability to accurately predict global climate and carbon-cycle response to forcing scenarios over the next 100 years. [1]
8. Model and understand the physics of the ice ages, including embedded abrupt climate change events such as the Younger Dryas, Heinrich, and Dansgaard-Oeschger events. [2]
9. Model and understand the key climate events in the early history of Earth and other planets. [3]

The committee believes that progress on the challenges marked [1] in this list (Challenges 1-3, 5, and 7) would immediately accelerate with advances in computing capability. Those marked [2] are using or will shortly use current capability computing. For the one marked [3], HECC will probably not play a big role within the next 5 years.

HECC in the atmospheric sciences mainly involves simulations based on the coupled multidimensional partial differential equations of fluid dynamics and heat and mass transfer. The fundamental atmospheric processes are driven by a variety of forces arising from radiation, moisture processes, chemical reactions, and interactions with land and sea surfaces. We need to increase the horizontal mesh resolution by a factor of between 4 and 10 to meet the rising demands from users of weather predictions. That, in turn, necessitates a hundred- to thousandfold increase in computing capability.

Climate simulations are improved largely by the addition of physical processes and the coupling of processes, both of which exacerbate the computational challenge. Compounding that difficulty is the need to compress longer simulated periods into no more than a few months of run time. Among climate scientists, a minimum of 10 simulated years of model integration is typically thought to be needed for credible work. Such a standard, on top of increasing complexity and a need for higher resolution, creates yet more demand for HECC.

Evolutionary Biology

The committee identified the following as the major challenges facing evolutionary biology today:

1. What has been the history of life?
2. How do species originate?
3. How has life diversified across space and time?
4. What determines the origin and evolution of the phenotype?
5. What are the evolutionary dynamics of the phenotype-environment interface?
6. What are the patterns and mechanisms of genome evolution?
7. What are the evolutionary dynamics of coevolving systems?

Observation and experiment continue to be productive modes of inquiry for these major challenges, and because computational evolutionary biology is still young, progress through computational research is still possible with modest computing capabilities. Today, researchers can still investigate many aspects of these major challenges—pose questions, explore relationships and models, and develop algorithms—without reaching the level of complexity that calls for HECC. But for many, that is rapidly changing. Resolving relationships among species, individuals, or genes or analyzing the huge amount of available genomic data are already driving many of these computational efforts toward algorithmic and computational complexity and, thus, the need for HECC resources. Some research into Major Challenges 1, 3, 6, and 7 is already making use of capability computing. Eventually, HECC will be necessary to make progress on all of these challenges, as evolutionary biology relies more heavily on data mining and modeling.

Evolutionary biology is somewhat of a special case because the rapid development of large-scale genomic analysis has enabled radical new approaches, and the field is very much in transition from one based on observation to one based on massive amounts of genomic data. The eventual impacts of HECC are clear and enormous, but the field is only beginning to exploit HECC.

Chemical Separations

The issues facing chemical separations are very different from those discussed in connection with the other fields because it is a field that is dominated by the industrial sector and uses well-understood twentieth-century technologies that would be expensive to reformulate. However, competition in the industry is exerting pressure on U.S. manufacturers to develop technologies for more energy-efficient and environmentally friendly separations processes to better manufacture pharmaceuticals, reduce greenhouse gases in emissions, and increase drinking water supplies that may become scarce in the future. It is very difficult to develop these demanding processes through experimentation alone, and so the time is coming when advanced computation will be more critical for technological advance.

For instance, distillation is highly effective for separating compounds based on differences in their relative volatilities. Yet, because it requires that the mixture be repeatedly vaporized and condensed, it consumes very large amounts of energy. Nonetheless, distillation is by far the most common separation process, used in as much as 80 percent of the most common chemical separations processes, so that optimization of phase equilibria will remain important for the chemical separations industry. Mass-separating agents (MSAs)—solvents, absorbents, adsorbents, membranes, and so on—are also now being used to amplify the separating capability and provide more economical and environmentally friendly solutions. However, the design of new MSA-based systems is severely hindered by the lack of physical property data. HECC has the potential to lead to significant breakthroughs in the development of new MSAs and, thus, markedly reduce the energy consumed by separation systems.

The following are the major challenges in chemical separations found by the committee:

1. How can we predict physical properties at the level of accuracy required for defining the optimal conditions for separating mixtures?
2. How can we design, construct, and produce MSAs with appropriately engineered three-dimensional structures (when needed) that facilitate the rapid and efficient accomplishment of difficult separations?
3. How can we design overall separation systems that incorporate several individual separation units for economically optimal separations of complex mixtures?

There are numerous examples of computational chemistry leading to new understanding of the behavior of chemicals in separation systems, and great potential for further benefits in addressing Major Challenges 1 and 2. Because simulations of molecules of industrial importance and of realistic systems are computationally demanding, it is likely that HECC resources will be required for those applications. Major Challenge 3 calls for the ability to optimize the interplay of multiple separation processes to achieve high-performance separation systems for complex mixtures. This is a very demanding computational task, but much work must be carried out before it can be addressed through HECC.

In summary, HECC can play a transformative role in chemical separations by directing experimentation in more productive directions. It can play that role by providing (1) more accurate phase equilibria data for a wider range of chemical compounds and multicomponent systems—and with greater safety when the separations deal with chemical species that may be toxic or dangerously reactive—and (2) fast screening of, and design information for, candidate MSAs. Ultimately, HECC holds the promise of enabling more optimal design of complete chemical separation processes.

CROSSCUTTING OBSERVATIONS

Chapter 6 gives an indication of the requirements in mathematics, computer science, and computing infrastructure associated with the technical challenges identified in Chapters 2 through 5. For astrophysics, only a small fraction of the algorithms of importance are able to scale well to 10^3 - 10^4 processors, to say nothing of scaling to even larger platforms. Algorithms, models, and software are therefore needed to enhance scalability for a large number of applications. New models are needed to represent multiscale physics in a way that maps well onto as-yet-undefined new computer architectures. Algorithms for discretization, solution of stiff systems of differential equations, and data management are also critical.

In the long term, the atmospheric sciences also will need new algorithms for discretization, solution of stiff systems of differential equations, and data management, and they will need new models to capture multiscale and multiphysics phenomena. However, for the near term, the field could readily exploit a

10-fold increase in computing capability to improve prediction of severe weather and climate prediction; to better support critical industries such as transportation, energy, and agriculture; and to increase the atmospheric grid resolution of a coupled atmosphere-ocean-sea-ice-biogeochemistry climate model.

Evolutionary biology is not currently limited by HECC resources, but that situation is changing rapidly. Adding more species and making more use of genomic data will quickly drive computational evolutionary biology into the realm of high-end computing. Indeed, scalability problems with many algorithms and the massive amounts of genomic data to be exploited will soon limit evolutionary biology if it does not get adequate HECC resources.

Progress in chemical separations is not currently limited by HECC, but important opportunities would open up from increasing computing capabilities. In the short term, the major HECC requirements are those that would enable simulations to be performed more readily so that their ability to guide experimentation could be exploited more routinely. At present, it is often just easier to perform experiments, because the simulations cannot offer enough predictive power. In the longer term, researchers would like to improve the accuracy of the underlying molecular models for a wider range of materials. An equally desirable capacity would be to enable the convergence of simulated observables so as to attain independent prediction of materials properties over the ranges of pressure and temperature needed for their phase diagrams. Achieving this capability will require algorithms and software that routinely use 10^5 - 10^6 processors per run, which will in turn require new ideas in mathematical models, numerical algorithms, and software infrastructure.

A common challenge facing three of the committee's fields (excepting chemical separations) is managing and exploiting massive (and increasing) amounts of data. Failure to address this issue—which increasingly requires capability computing—could limit our nation's ability to profit from past and ongoing investments in observation and experiment.

Another common challenge for all four fields is preparing the next generation of researchers, who will push the frontiers of computational science and engineering. Investment in education and training for computational science is needed in all core science disciplines where HECC currently plays, or will play, a larger role in meeting the major challenges. HECC investments should address needs for education and training infrastructure. Students need stronger foundations in mathematics and statistics. Two options for advancing the computational capabilities of our future workforce are (1) imparting stronger computational science skills through the normal curricula of HECC-dependent disciplines and (2) developing a distinct undergraduate or graduate track that produces computational “technologists,” professionals whose contribution to the research enterprise is enabling efficient and effective computational approaches rather than personally conducting the research. The first option would impose an additional burden on graduate students, forcing them to make curriculum trade-offs or take more time to complete a degree. The second option requires the definition of career optimizing tracks for computational generalists so they can move smoothly into and out of science domains as the need arises for their expertise, as well as cultural adjustments so that their contributions are recognized and career paths exist.

CONCLUSIONS

The committee members, in spite of their diverse backgrounds and varying degrees of reliance on high-end computing, readily agreed that HECC requires the integration of synergistic elements and should be managed as a system.

Conclusion 1. *High-end capability (HECC) computing is advanced computing that pushes the bounds of what is computationally feasible. Because it requires a system of interdependent com-*

ponents and because the mix of critical-path elements varies from field to field, HECC should not be defined simply by the type of computing platform being used. It is nonroutine in the sense that it requires innovation and poses technology risks in addition to the risks normally associated with any research endeavor.

High-end computational capabilities include whatever mix of hardware, models, algorithms, software, intellectual capacity, and computational infrastructure must be deployed to enable the desired computations. High-end computing platforms are certainly part of that mix, and the most ambitious and progressive computational science may, in many cases, require a new generation of hardware.

In Chapter 7 the committee lists 10 prerequisites if a field is to profit from HECC. These will be useful in evaluating investment opportunities.

Conclusion 2. *Advanced computational science and engineering is a complex enterprise that requires models, algorithms, software, hardware, facilities, education and training, and a community of researchers attuned to its special needs. Computational capabilities in different fields of science and engineering are limited in different ways, and each field will require a different set of investments before it can use HECC to overcome the field's major challenges.*

At the very least, HECC infrastructure will consist of hardware, operating software, and applications software. In addition, there will continue to be a need for data management tools, graphical interface tools, data analysis tools, and algorithms research and development. Disciplines will take advantage of the increased availability of HECC in proportion to how much of the necessary infrastructure has already been created—that is, whether the field has achieved a state of readiness.

Conclusion 3. *Decisions about when, and how, to invest in HECC should be driven by the potential for those investments to enable or accelerate progress on the major challenges in one or more fields of science and engineering.*

Once a decision is made to invest in computational resources, that investment must provide all the elements of infrastructure that are needed by the fields likely to use the resource.

Conclusion 4. *Because the major challenges of any field of science or engineering are by definition critical to the progress of the field, underinvestment in any of them will hold back the field.*

Optimum progress across all the major challenges will be achieved if all modalities of research—*theoretical, experimental, and computational*—are supported in a balanced way.

In many cases, HECC capabilities must continue to be advanced to maximize the value of data already collected or investments already made in experimental facilities. For instance, remote sensing projects under way in astrophysics and the atmospheric sciences will produce quantities of data that cannot be utilized by those fields without commensurate progress in analytical capabilities.

Conclusion 5. *The emergence of new hardware architectures precludes the option of just waiting for faster machines and then porting existing codes to them. The algorithms and software in those codes must be reworked.*

There do not yet exist productive and easy-to-use programming methodologies or low-level blocks

of code that can take full advantage of multicore processors. Multicore parallelism is unfamiliar to many commercial software developers, and it also requires different sorts of parallel algorithm development.

Conclusion 6. *All four fields will need new, well-posed mathematical models to enable HECC approaches to their major challenges. Astrophysics and the atmospheric sciences share two needs: one for new ways to handle stiff differential equations and one for continuing advances in multiresolution and adaptive discretization methods. Astrophysics and chemical separations also share two needs: one for accurate and efficient methods for evaluating long-range potentials that scale to large numbers of particles and processors and one for stiff integration methods for large systems of particles.*

In addition, it is clear to the committee that the management, analysis, and mining of data present an increasingly critical and crosscutting algorithmic challenge. Enormous sets of input data, such as those from satellites and telescopes, require HECC to digest data and elicit insights.

Conclusion 7. *To capitalize on HECC's promise for overcoming the major challenges in many fields, there is a need for students in those fields, graduate and undergraduate, who can contribute to HECC-enabled research and for more researchers with strong skills in HECC.*

The committee foresees a growing need for computational scientists and engineers who can work with mathematicians and computer scientists to develop next-generation HECC software. Chapters 4, 5, and 6 explicitly mention a need for the more widespread teaching of scientific computing. Specifying an optimal career path for people who are able to straddle HECC and a traditional discipline is problematic, especially in academia. What is needed is a path that encompasses both a service role (HECC consulting within their field and to computer scientists) and opportunities to conduct their own research.

Even though the four fields selected for this study are disparate, the committee was able to develop major challenges for each and then determine which of those challenges are critically dependent on HECC. The following are suggestions for evaluating the potential impacts of HECC in other fields:

- It is necessary to build on the existing consensus about a field's current frontiers or major challenges. Developing from scratch a consensus picture of the frontier and of the major challenges that define promising directions for extending that frontier is in itself a sizable task.
- It is important to determine which major challenges for the field are critically dependent on HECC. While it is easy to spot opportunities for applying HECC to gain advantage, that is not the same as identifying the challenges whose progress will be impeded without the use of appropriate HECC.

All the infrastructure components needed to apply HECC to the challenges that depend on it must be identified, and the community must develop a clear understanding of the resources needed to complete the infrastructure. Merely giving a field access to supercomputers is no guarantee that the field's scientific progress will be enabled or accelerated.

THE POTENTIAL IMPACT OF
**HIGH-END CAPABILITY
COMPUTING**
ON FOUR ILLUSTRATIVE FIELDS OF
SCIENCE AND ENGINEERING

Committee on the Potential Impact of High-End Computing on
Illustrative Fields of Science and Engineering

Division on Engineering and Physical Sciences

Division on Earth and Life Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS

500 Fifth Street, N.W.

Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. NCO-0610176 between the National Academy of Sciences and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Library of Congress Cataloging-in-Publication Data

National Research Council (U.S.). Committee on the Potential Impact of High-End Computing on Illustrative Fields of Science and Engineering

The potential impact of high-end capability computing on four illustrative fields of science and engineering / Committee on the Potential Impact of High-End Computing on Illustrative Fields of Science and Engineering, Division on Engineering and Physical Sciences, Division on Earth and Life Sciences.

p. cm.

Includes bibliographical references.

ISBN 978-0-309-12485-0 (pbk.) — ISBN 978-0-309-12486-7 (pdf) 1. High performance computing—United States. 2. Supercomputers—United States. 3. Science—Data processing. 4. Engineering—Data processing. I. National Research Council (U.S.). Division on Engineering and Physical Sciences. II. National Research Council (U.S.). Division on Earth and Life Studies. III. Title.

QA76.88.N35 2008

502.85—dc22

2008037012

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

Copyright 2008 by the National Academy of Sciences. All rights reserved.

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. Ralph J. Cicerone 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. Charles M. Vest 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. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

**COMMITTEE ON THE POTENTIAL IMPACT OF HIGH-END COMPUTING ON
ILLUSTRATIVE FIELDS OF SCIENCE AND ENGINEERING**

JOHN W. LYONS, National Defense University, *Chair*
DAVID ARNETT, University of Arizona
ALOK N. CHOUDHARY, Northwestern University
PHILLIP COLELLA, Lawrence Berkeley National Laboratory
JOEL L. CRACRAFT, American Museum of Natural History
JOHN A. DUTTON, Storm Exchange, Inc., and Pennsylvania State University (retired)
SCOTT V. EDWARDS, Harvard University
DAVID J. ERICKSON III, Oak Ridge National Laboratory
TERESA L. HEAD-GORDON, University of California at Berkeley
LARS E. HERNQUIST, Harvard-Smithsonian Center for Astrophysics
GEORGE E. KELLER II, MATRIC
NIPAM H. PATEL, University of California at Berkeley
MARY E. REZAC, Kansas State University
RONALD B. SMITH, Yale University
JAMES M. STONE, Princeton University
JOHN C. WOOLEY, University of California at San Diego

Staff

SCOTT WEIDMAN, Study Director, Division on Engineering and Physical Sciences
BARBARA WRIGHT, Administrative Assistant

Preface

The study that led to this report was called for in the President's Fiscal Year 2006 Budget. In accordance with that call, the study was commissioned by the National Coordination Office (NCO) for Networking and Information Technology Research and Development (NITRD), which supports the NITRD Subcommittee, which operates under the White House National Science and Technology Council to coordinate federal investments in networking and information technology research and development. NCO efforts were guided by a steering group appointed by the NITRD Subcommittee and input from the White House Office of Science and Technology Policy and the White House Office of Management and Budget. The study was conducted by a committee (see Appendix A) constituted under the National Research Council's (NRC's) Division on Engineering and Physical Sciences and its Division on Earth and Life Sciences.

This report addresses the following charge:

The study will develop a better understanding of the potential scientific and technological impact of high-end capability computing in four illustrative fields of S&E of interest to the federal government. More specifically, the study will

- (a) Review the important scientific questions and technological problems identified for those fields in other sources (e.g., decadal surveys);
- (b) Identify the subset of those important questions and problems for which an extraordinary advancement in understanding is difficult or impossible without high-end capability computing;
- (c) Identify some of the likely impacts of making progress on as many of the scientific questions and technological problems identified in (b) as possible and the contribution that high-end capability computing can make to this progress;
- (d) Discuss some of the most significant ramifications of postponing this use of high-end capability computing in order to capitalize on the decreasing cost of computing over time;
- (e) Identify the numerical and algorithmic characteristics of the high-end capability computing requirements needed to address the scientific questions and technological problems identified in (b); and

- (f) Categorize the numerical and algorithmic characteristics, specifically noting those categories that cut across disciplines. This task shall be done in a way that can later be used to inform design and procurements of high-end capability systems.

This list of tasks is not in priority order. Tasks (a), (b), (e), and (f) are considered to be the most important and essential for the study's success.

When the NCO asked the NRC to undertake this study, it requested that the committee consider, for specificity, the potential impacts of high-end capability computing (HECC) on four illustrative fields chosen from the following broad areas of science and engineering that are of importance to the federal government: physics and/or astronomy, the geosciences, chemistry and/or chemical engineering, and the biological sciences and/or biomedical engineering.

NRC staff from the Division on Engineering and Physical Sciences and the Division on Earth and Life Sciences developed the following criteria for selecting illustrative fields from within those broad areas:

- Each field selected for study should have relatively well-defined goals, and the goals should be documented in consensus reports so that the current study need not identify the major research challenges from scratch.
- The fields selected should be broad enough to be of widespread interest yet limited enough to be fairly represented through the study's process, which is outlined below.
- Collectively, the four fields selected should span the most important areas of interest to the NITRD community: the geosciences, defense, health, space, energy, commerce, and basic research.
- The four fields selected should illustrate a range of comfort, usage, and acceptance of HECC approaches to investigation.

By meeting these four criteria, it was felt that the fields selected would adequately illustrate the range of science and engineering research topics that the NITRD program encompasses and most likely raise a broad set of HECC-related issues that the NITRD program addresses. The committee did not believe it was necessary, or even desirable, to exhaustively survey the entire community supported by NITRD's HECC programs, nor was that tactic called for in the charge.

Based on these criteria, NRC staff held a number of discussions with experts in physics, astronomy, the geosciences, chemistry, chemical engineering, and biology and with representatives of the NITRD community. An attempt was made to include at least one field that is not obviously dependent on HECC. As a result, it was decided to focus the study on the following four illustrative fields of science and engineering:

- Astrophysics,
- The atmospheric sciences,
- Evolutionary biology, and
- Chemical separations.

While this study does identify the potential impact of HECC in these four fields, and thus implicitly points out some funding opportunities, that is not its real goal, and this study is no substitute for competitive review of specific proposals. Rather, the study is meant to define the sort of examination that any field or federal agency could undertake in order to analyze the HECC infrastructure it needs to support progress against its research goals, within the context of other means of attacking those goals.

In the course of such an examination, a field or agency will methodically evaluate its ability to take advantage of HECC infrastructure and in so doing will identify the components of the infrastructure that are needed to enable progress. By defining the self-examination, the study also implies a vision for the environment that will best profit from HECC investments, and it provides guidance for federal policy makers who must weigh competing requests for those funds.

The committee met four times in the course of its study; agendas are included in Appendix B. Of special note was the second meeting, which included four small workshops in parallel (one for each of the four fields) to obtain community input on Tasks (a)-(e) of the study's charge. The list of participants in those workshops is also in Appendix B.

The goal of these information-gathering workshops was not to canvass broadly—the study was not meant to identify *de novo* the leading questions in the four fields and relied on other sources—but to have a focused discussion with a small number of leaders in each field plus federal scientists and engineers. The workshops discussed whether a draft list of major research challenges adapted from other sources would be adequate for the purposes of this study; identified which of those challenges are critically dependent on HECC, and in what way(s); evaluated the abilities of the four fields to capitalize on HECC for addressing those challenges; and identified critical gaps in each field's capabilities for realizing the opportunities presented by HECC. The results of these discussions are captured in Chapters 2-5.

The committee's third and fourth meetings focused on crosscutting elements of the charge, primarily Task (f). Those discussions led to Chapter 6, which presents and analyzes the HECC capabilities and needs for the four fields, and Chapter 7, which essentially provides a decision process for fields, agencies, and federal policy makers to use in assessing the potential value of HECC investments.

Acknowledgments

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 Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its 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:

THOM H. DUNNING, University of Illinois at Urbana-Champaign,
JOSEPH FELSENSTEIN, University of Washington,
JAMES HACK, Oak Ridge National Laboratory,
DAVID M. HILLIS, University of Texas-Austin,
JOHN P. HUCHRA, Harvard-Smithsonian Center for Astrophysics,
ROBERT F. LUCAS, University of Southern California,
CHRISTOPHER MCKEE, University of California at Berkeley,
DANIEL A. REED, Microsoft Corporation,
RICHARD B. ROOD, University of Michigan, and
JEFFREY SIIROLA, Eastman Chemical Company.

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 Margaret H. Wright of New York University. Appointed by the National Research Council, she 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.

The committee also acknowledges the valuable contribution of the following individuals, who provided input at the meetings on which this report is based:

TOM ABEL, Stanford University,
ARDEN BEMENT, National Science Foundation (NSF),
JOAN BRENNECKE, University of Notre Dame,
ANTONIO J. BUSALACCHI, University of Maryland,
ANNE CHAKA, National Institute of Standards and Technology,
DANIEL DRELL, Department of Energy,
SEAN EDDY, Howard Hughes Medical Institute,
SERGEY GAVRILETS, University of Tennessee, Knoxville,
BRIAN GROSS, NOAA Geophysical Fluid Dynamics Laboratory,
JAMES HACK, Oak Ridge National Laboratory,
SALLY HOWE, National Coordination Office for Networking and Information Technology
Research and Development
JOEL KINGSOLVER, University of North Carolina,
JOHN MARBURGER, Office of Science and Technology Policy,
CHRISTOPHER McKEE, University of California at Berkeley,
EVE OSTRIKER, University of Maryland,
JOEL PARRIOTT, Office of Management and Budget,
DANIEL ROKHSAR, Lawrence Berkeley National Laboratory,
ED SEIDEL, Louisiana State University,
NIGEL SHARP, NSF,
JEFFREY SIIROLA, Eastman Chemical Company,
RICK STEVENS, Argonne National Laboratory,
GEORGE STRAWN, NSF,
ALEX SZALAY, Johns Hopkins University,
SIMON SZYKMAN, National Coordination Office for Networking and Information Technology
Research and Development
CHAOWEI YANG, NASA Applied Sciences Program, and
MANFRED ZORN, NSF.

The committee is also grateful for the contributions of James McGee and Ann Reid of the National Academies staff, who assisted in the early stages of this study.

Contents

SUMMARY	1
1 PROBLEM DEFINITION AND HISTORY	9
Introduction, 9	
History of High-End Computing, 11	
Current State of High-End Capability Computing, 13	
2 THE POTENTIAL IMPACT OF HECC IN ASTROPHYSICS	15
Introduction, 15	
Major Challenges in Astrophysics, 17	
Major Challenges That Require HECC, 18	
Methods and Algorithms in Astrophysics, 30	
HECC for Data Analysis, 32	
Realizing the Potential Impact of HECC on Astrophysics, 33	
References, 35	
3 THE POTENTIAL IMPACT OF HECC IN THE ATMOSPHERIC SCIENCES	37
Introduction, 37	
Major Challenges in the Atmospheric Sciences, 39	
Computational Challenges in the Atmospheric Sciences, 46	
The Need for HECC Resources to Advance the Atmospheric Sciences, 57	
Conclusion: Earth in a Computer, 60	
References, 60	
4 THE POTENTIAL IMPACT OF HECC IN EVOLUTIONARY BIOLOGY	63
Introduction, 63	
Major Challenges of Evolutionary Biology, 64	

	Major Challenges in Evolutionary Biology That Require HECC, 79	
	References, 85	
5	THE POTENTIAL IMPACT OF HECC IN CHEMICAL SEPARATIONS	89
	Introduction, 89	
	Major Challenges Facing Chemical Separations, 92	
	Potential Impacts of HECC for Chemical Separations, 96	
	Current Frontiers of HECC for Chemical Separations, 99	
	Other Issues That Limit the Value of HECC to Chemical Separations, 103	
	References, 103	
6	NUMERICAL AND ALGORITHMIC CHARACTERISTICS OF HECC THAT WILL BE REQUIRED BY THE SELECTED FIELDS	105
	Numerical and Algorithmic Characteristics of HECC for Astrophysics, 105	
	Numerical and Algorithmic Characteristics of HECC for the Atmospheric Sciences, 108	
	Numerical and Algorithmic Characteristics of HECC for Evolutionary Biology, 110	
	Numerical and Algorithmic Characteristics of HECC for Chemical Separations, 112	
	Categorization of Numerical and Algorithmic Characteristics of HECC Needed in the Four Selected Fields, 114	
	Crosscutting Challenges from Massive Amounts of Data, 116	
	Crosscutting Challenges Related to Education and Training, 118	
7	CONCLUSIONS	121
	Supporting High-End Computational Research, 121	
	The Need for Continuing Investment in HECC, 123	
	Classes of Numerical and Algorithmic Challenges, 125	
	Human Resources, 125	
	Lessons Learned for Fields That Might Perform Similar Studies, 126	
APPENDIXES		
A	Biographical Sketches of Committee Members	129
B	Agendas of Committee Meetings	133
C	Glossary	141

THE POTENTIAL IMPACT OF
**HIGH-END CAPABILITY
COMPUTING**
ON FOUR ILLUSTRATIVE FIELDS OF
SCIENCE AND ENGINEERING

