

## 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](http://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