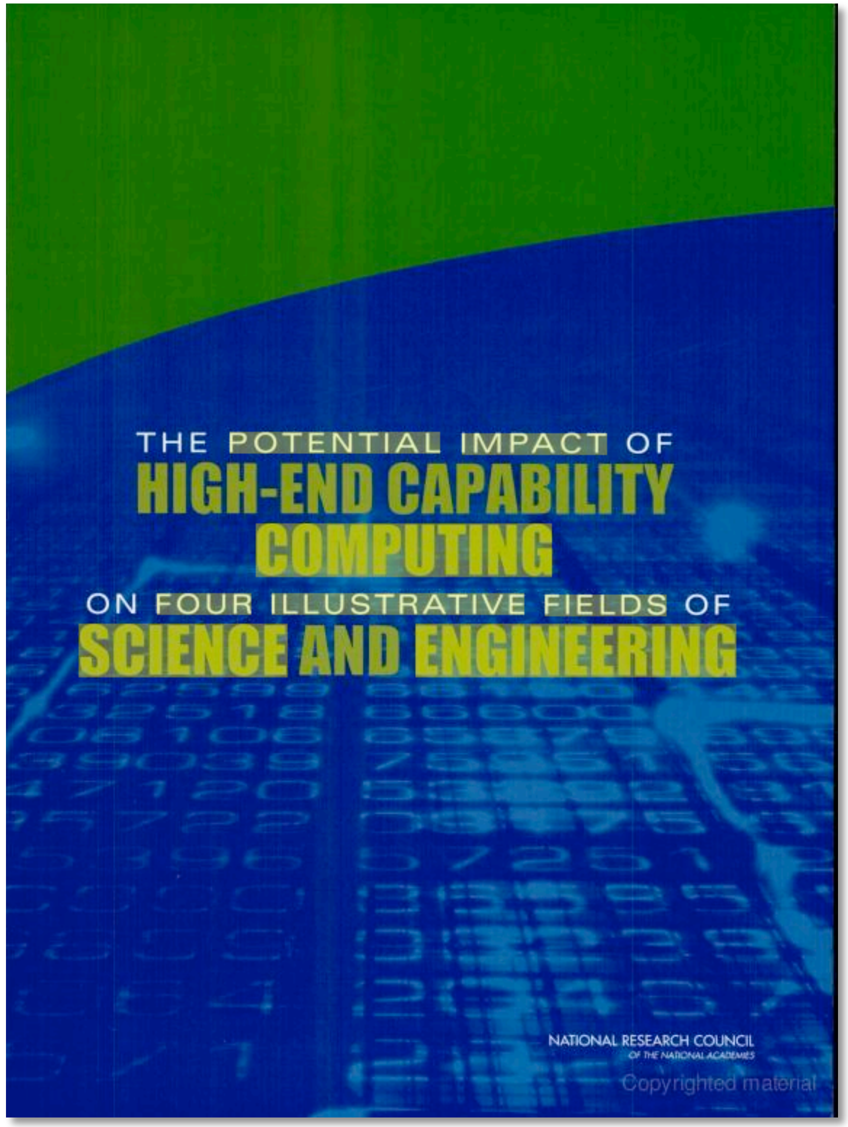
THE POTENTIAL IMPACT OF
HIGH-END CAPABILITY
COMPUTING
ON FOUR ILLUSTRATIVE FIELDS OF
SCIENCE AND ENGINEERING
NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES
Copyrighted material

Study 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.

Four Fields

- Astrophysics
- Atmospheric Science
- Chemical Separations
- Evolutionary Biology

Study Committee

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 Penn. State Univ. (ret.)
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

NRC Study Director: Scott Weidman
Sponsor: NITRD

Study Approach

- Four meetings.
- Small disciplinary workshops in December of 2006.
- External review.
- Official release on August 26, 2008.
- Related symposium at NAS on September 22, 2008.
- Briefed study to OMB, OSTP on November 5, 2008.

Conclusions (Excerpt)

Conclusion 1. **High-end capability computing is advanced computing that pushes the bounds of what is computationally feasible.** Because it requires a system of interdependent components and the mix of critical-path elements varies from field to field, it 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.

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.

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.**

Astrophysics – Major Challenges
Those requiring HECC are in boldface

- What is dark matter?
- What is dark energy?
- **How did galaxies, quasars, and black holes form?**
- **How do stars and planets form? And evolve?**
- **What are the mechanisms for supernovae and gamma ray bursts?**
- **What will the universe look like observed in gravitational waves?**

Atmospheric Sciences – Major Challenges

- **Extend the range, accuracy and utility of weather prediction.**
- **Improve understanding and prediction of severe weather, pollution, and climate events.**
- **Effect of seasonal, decadal, and century-scale climate variation on global, regional, and local scales.**
- Understand the physics and dynamics of clouds, aerosols, and precipitation.
- **Understand effects of moisture and chemical exchange at Earth's surface.**
- Develop theory for nonlinearities and tipping points in weather and climate systems.
- **Create ability to predict global change over the next 100 years.**
- Understand the physics of the ice ages, including abrupt climate change.
- Understand key climate events in the early history of Earth and other planets.

Evolutionary Biology – Major Challenges

- **Understand the history of life.**
- **Understand how species originate.**
- Understand diversification of life across space and time.
- **Understand the origin and evolution of the phenotype.**
- Understand the evolutionary dynamics of the phenotype-environment interface.
- Understand the patterns and mechanisms of genome evolution.
- Understand the evolutionary dynamics of coevolving systems.

Chemical Separations – Major Challenges

- **Predict physical properties for phase equilibria (for difficult separations).**
- **Design and produce mass separating agents.**
- Design optimal separation systems with multiple separation units.

Crosscutting Issues

The Charge

“Identify the numerical and algorithmic characteristics of the HECC requirements needed to address the scientific questions and technological problems indentified in [Chapters 2-5]”.

“Categorize the numerical and algorithmic characteristics, specifically noting those requirements that cut across disciplines”.

Approach to responding to this part of the charge: analysis of the requirements for each science / technology area, followed by discussion of crosscuts.

Potential Rate-Limiting Issues

- Models
- Algorithms
- Software Infrastructure
- Facilities
- Data Analysis and Management
- People

Astrophysics

•Transition from well-characterized “first-principles” models to multiscale / multiphysics models.

• Algorithmic issues: mulitresolution methods, particularly for multiphysics coupling (e.g. radiation hydrodynamics, general relativistic fluid dynamics); implicit methods for stiff time scales.

• Big data, from both observations and simulation.

Atmospheric Sciences

• Near-term need for 10X increase in capability, with concomitant effort in scaling codes to 10^4 processors (particularly for Numerical Weather Prediction).

• The next major increment in model fidelity for climate could be obtained by a transition from statistical models of clouds / precipitation (valid down to 25 km) to cloud-system resolving models requiring 1 km resolution in the tropics. Leads to a major reconsideration of many components of the model, algorithm and software space.

• New opportunities / requirements in data assimilation due to the vast increase in sources of data (NWP) or the application of data assimilation to climate modeling.

Evolutionary Biology

Current use of computing is mainly discrete mathematics and statistics for data mining. These areas are transitioning from workstation-based activities to HECC.

Modeling is in its infancy

- Enormous range of spatial scales (from cellular level to populations) and temporal scales (seconds up to geological time scales).
- No first-principles models.

Chemical Separations

- Current state of the art in computational chemistry adequate for investigation of qualitative behavior of separations processes, but not for end-to-end quantitative prediction and optimization.

- Severe tradeoffs between fidelity and computability – combination of model and algorithmic issues.

- Focus on process optimization requires new mathematical and algorithmic tools.

- Workforce issues a concern.

Crosscutting Issues - Simulation

- Models. (Further) movement away from first-principles models to ones that are bootstrapped from experimental / observational data and more detailed auxiliary computer simulations. More of an emphasis on hybrid stochastic / deterministic models.

- Algorithms. Overlapping requirements include multiresolution methods, methods for stiff systems, and high-performance particle methods. Strong interaction with new models.

- Software.
  - End of Moore's law (10^{8}-10^{9} threads), complexity of multiscale, multiphysics simulation codes are prominent concerns here, with major impacts on software productivity.
  - All four fields are willing to adopt to various extents the use of shared software infrastructure. Tradeoffs between domain-specificity and shared software are different for different fields. Where will the software come from?

Crosscutting Issues - Data

- Astrophysics, atmospheric sciences and evolutionary biology are data-intensive.
  - High-thoughput experiments / observations.
  - Simulation data.
- Challenges:
  - Knowledge discovery from data
  - Sharing of data
  - Analysis / mining of data
  - High-performance input / output
- Strong variation in the interaction of data with simulation.

Crosscutting Issues – Education and Training

- All four fields have, to varying degrees, been successful in integrating computational science into their disciplines.

- Need for better training in the fundamental mathematics and computer science that underpins the use of HECC.

- Greater emphasis on software development in education of computational scientists / engineers. Most Ph. D. thesis projects are "proof-of-principle" implementations, with no persistent software artifacts.

- Two approaches to integration of CSE into science disciplines: disciplines take ownership, or form a new CSE disciplinary area.

- Career track for applications software developers.

Conclusions (cont.)

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.

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.**

Conclusion 6. **All four fields will need new, well-posed mathematical models to enable HECC approaches to their major challenges.** Astrophysics and atmospheric science 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.

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.**

Final Comments

- Operational realities of selling science programs to policymakers. A airtight detailed science case, placed in larger context; timing issues (“Why this … Why now ?”).
- Cheerleading by computational scientists is not credible – need outside validation.
- Models (to some extent), algorithms, facilities all have well-established funding mechanisms, and big data is getting a strong push. What about software ?
- CSE ecosystem: training, career tracks, persistent institutional commitments.